

Computationally Efficient Deniable Communication

Qiaosheng Zhang

Mayank Bakshi

Sidharth Jaggi

zq015@ie.cuhk.edu.hk

mayank@inc.cuhk.edu.hk

jaggi@ie.cuhk.edu.hk

Institute of Network Coding, Chinese University of Hong Kong

Abstract

In this paper, we design the first *computationally efficient* codes for simultaneously *reliable* and *deniable* communication over a Binary Symmetric Channel (BSC). Our setting is as follows – a transmitter Alice wishes to potentially reliably transmit a message to a receiver Bob, while ensuring that the transmission taking place is deniable from eavesdropper Willie (who hears Alice’s transmission over a *noisier* BSC). Prior works show that Alice can reliably and deniably transmit $\mathcal{O}(\sqrt{n})$ bits over n channel uses without any shared secret between Alice and Bob. One drawback of prior works is that the computational complexity of the codes designed scales as $2^{\Theta(\sqrt{n})}$. In this work we provide the first computationally tractable codes with provable guarantees on both reliability and deniability, while simultaneously achieving the best known throughput for the problem.

I. INTRODUCTION

Alice may or may not wish to communicate with a receiver Bob over a Binary Symmetric Channel with crossover probability p_b , denoted by $\text{BSC}(p_b)$. However, an adversary Willie is able to eavesdrop on their communication over a “noisier” Binary Symmetric Channel – $\text{BSC}(p_w)$ (here p_w is strictly larger than p_b ¹), and only cares about whether Alice is transmitting or not. Therefore, Alice would like to use a novel communication scheme to prevent her transmission status from being detected by Willie (*deniable* from Willie) and also ensure that her messages are received by Bob correctly.²

We first give an overview of several problems related to our setup. Shannon first defined the concept of *information-theoretic security* [2], which requires the key rate to be as large as the message rate to achieve perfect secrecy. *Kerckhoff’s principle* [3], however, states that a system should be secure even if everything about the system, except the key, is public knowledge. Wyner demonstrated that shared secrets can be replaced with asymmetry in channel noise [4], [5] (as in this work). The reader is referred to [6], [7] for recent surveys on physical-layer security. The classical *steganography* problem, which considers how to hide a undetectable message in plain sight, has been well-studied – see, for instance, [8] for a survey. Cachin [9] first focused on the problem of information-theoretic steganography, and Maurer [10] drew connections between the problem of steganography and that of *hypothesis testing*. In [11], Wang and Moulin gave an information-theoretic characterization of the capacity of the perfectly secure steganography problem (with unbounded-sized shared secrets between Alice and Bob).

We now turn to reliable and deniable communication, which is the main focus of this work. As alternatives to deniability, people also use different terms such as *covert*, *stealthy* and *low probability of detection* (LPD) in the literature, to define essentially the same security requirement. Bash *et al.* gave the first results on information-theoretically guaranteed LPD communication over noisy AWGN channels [12]–[15]. Noting that the result of Bash *et al.* relied critically on the presence of large shared secrets between Alice and Bob³, Che *et al.* designed reliable and deniable (and information-theoretically secure) communication schemes over BSCs without using *any* shared secrets, relying only on the asymmetry of level of channel noise on the two channels [16]–[19]. The work of [20] studied stealthy communication from a channel resolvability approach, while Wang *et al.* [21] and Bloch [1] first derived tight capacity characterizations for discrete memoryless channels (DMCs). We discuss the intuition behind these schemes in greater detail in Section II below.

While the plethora of codes and bounds in the recent literature paint a clear picture of the limits of reliable communication possible between the transmitter Alice and the receiver Bob while remaining deniable (or stealthy/covert/LPD) from the eavesdropper Willie, prior to this work there were still no computationally efficient communication schemes with information-theoretic proofs of deniability. Though a variety of computationally-efficient schemes [22]–[27] give good heuristics for such communication, they typically lack proofs that the proposed schemes do indeed provide information-theoretic deniability of such detectors that may be employed by the eavesdropper, regardless of the computational complexity.

In this paper, we present the first coding scheme which has provable throughput and deniability guarantees while ensuring that the computational complexity for both encoding and decoding is at most polynomial in the number of transmitted message bits. The rest of this paper is organized as follows. We formally describe our model in Section III. In Section IV, we give the

The work of Qiaosheng Zhang, Mayank Bakshi and Sidharth Jaggi described in this paper was partially supported by a grant from University Grants Committee of the Hong Kong Special Administrative Region, China (Project No. AoE/E-02/08).

¹Note that without this asymmetry, whenever Bob can decode reliably, so can Willie.

²For ease of exposition, in this work we focus on scenarios in which all channels are BSCs. However, following the lead of [1], it is likely that these results can be directly generalized to other DMCs.

³In fact, the size of the keys required by their scheme is larger than the throughput from Alice to Bob.

main result of this paper, a performance characterization of a specific class of computationally-efficient reliable and deniable communication schemes, and describe the corresponding codes in Sections V. Section VII contains mathematical preliminaries, and Sections VIII and IX provide the proofs of deniability and reliability respectively of our codes.

II. INTUITION

We begin by first giving an intuitive description of our work and place it in the context of prior works.

A. Challenges

The intuition behind the deniable schemes first presented in [12] and elaborated on in other works such as [1], [13], [15]–[19], [21], [28] is that most reasonable noise processes have, with non-zero probability, “some deviation” in the “noise intensity”. For instance, a length- n Bernoulli(p_w) sequence (corresponding to the additive noise sequence in a BSC(p_w) – a *Binary Symmetric Channel with crossover probability* p_w – the channel from the transmitter Alice to the eavesdropper Willie) has expected value np_w , but has standard deviation $\sqrt{np_w(1-p_w)}$. Hence, if Alice uses a carefully designed codebook containing codewords with low Hamming weight (about $\mathcal{O}(\sqrt{n})$) then the expected “power density” at the eavesdropper (about $np_w + \mathcal{O}(\sqrt{n})$) may reasonably be attributed by Willie to natural variations in the noise-level he observes. Further, it is also known [16] that to ensure deniability in communication, one *must* use codes with very low average Hamming weight (*i.e.* with weights no larger than⁴ $\mathcal{O}(\sqrt{n})$). This restriction on codeword weights due to deniability from Willie, along with the requirement that Bob be able to reliably decode, implies that the optimal reliable throughput from Alice to Bob that is simultaneously deniable from Willie scales only as a factor of \sqrt{n} , rather than linearly in the number of channel uses (as is the common paradigm in Shannon theory). Hence the capacity of such deniable communication schemes converges to zero! The interesting “first-order” question, therefore, is how many bits can be communicated reliably (to Bob) and deniably (from Willie) as a function of the square-root of the number of channel uses.

However, just choosing a codebook with low average Hamming weight does not suffice in guaranteeing deniability. For instance, suppose Alice chooses a codebook containing length- n binary vectors such that about half of the first \sqrt{n} locations are non-zero, but *all* the succeeding $n - \sqrt{n}$ bits in each codeword *must* be zero. While such a codebook would satisfy the low average Hamming weight requirement, it is nonetheless still easy for Willie to detect whether or not Alice is transmitting in such a scenario. If Alice is silent, he would expect to see about $p_w\sqrt{n}$ non-zero values in the first \sqrt{n} locations of his observation (with a standard deviation of about $\mathcal{O}(n^{1/4})$), whereas if Alice were transmitting a non-zero codeword, he would expect to see about $n/2$ non-zero values in the same locations (again with a standard deviation of about $\mathcal{O}(n^{1/4})$). By relatively standard analysis from the hypothesis-testing literature [29], it can be shown his estimate of Alice’s transmission status would be correct with high probability (over the noise in the channel to him). Hence one needs “good spreading” of the bits in the codewords as well – not all codewords can have their support concentrated in the same small set of locations.⁵

While the above serves as good intuition for constructing deniable communication schemes, providing mathematical guarantees for a given code can be extremely challenging – one has to prove that two different probability distributions supported on an exponentially large set are “very close”. Specifically, one distribution, denoted p_0 , corresponds to the scenario when Alice is silent, and corresponds to a Binomial(n, p_w) distribution. The other, denoted p_1 , corresponds to the scenario when Alice is transmitting using some code \mathcal{C} . Both these distributions are supported on the set (of exponential size in the block-length n) of possible observations seen by the eavesdropper Willie. Since the structure of p_1 depends intimately on the structure of \mathcal{C} , characterizing the difference between p_0 and p_1 for any specific code, or specific ensembles of codes, can be quite complicated.

A second challenge is due to the fact that most computationally efficient code designs in the literature (see, for example, [30], [31]) naturally lead to codes such that the average Hamming weight of codewords in the code is tightly concentrated around half the block-length, $n/2$. As noted above, simply designing codes of block-length about $\mathcal{O}(\sqrt{n})$ and embedding the codewords into a pre-specified and publicly known set of about $\mathcal{O}(\sqrt{n})$ locations in length- n vectors padded with 0s *also* does not work. To the best of our knowledge, prior to this work there were no *constant composition codes* [32] with such low Hamming weight, with good spreading properties, that enable communication at rates close to the optimal rates characterized in [1], [16], [21], and that are simultaneously computationally-efficient to encode and decode.

B. Our approach

Our approach is to use *concatenated-style* codes, that are inspired by Forney’s classical work [33] that gave the first computationally-efficient codes for arbitrary channels that also approached capacity. Forney noticed that since the computational cost of Shannon’s random codes is exponential with the blocklength n , dividing the message into $\Theta(\log n)$ -sized chunks and applying Shannon’s codes on each chunk would ensure that the overall complexity is only polynomial with the total blocklength, while still operating at rates close to the channel capacity. However, naïvely applying this “divide-and-conquer” idea would

⁴The results of [18] indicate an interesting phenomenon when there is uncertainty about the *level of noise of the channel*, and the coherence time is “long” – then, in fact, the throughput can be shown to scale linearly with the number of channel uses, rather than as \sqrt{n} .

⁵Indeed, this is the intuition in some recent heuristic approaches [22]–[27] to designing covert communication schemes – codes designed via “spread spectrum” techniques are analyzed. However, an information-theoretically rigorous proof of the deniability of such schemes is lacking.

lead to an overall high decoding error probability owing to the small blocklength (and hence, relatively large decoding error probability) for each chunk and the large number of chunks. In order to overcome this, Forney’s solution was to combine the “inner code” provided by Shannon with an “outer code”. The purpose of the outer code – typically a Reed-Solomon (RS) code – is to computationally efficiently correct any chunks that are in error by paying a negligible rate penalty.

We follow Forney’s lead, but adapt our construction to the constraints imposed by deniability. Foremost, while Forney’s construction operates with $\Theta(n)$ message bits, in our setting, at most $\mathcal{O}(\sqrt{n})$ bits of reliable transmission are possible. Thus, to ensure that each chunk contains $\Theta(\log n)$ message bits, the blocklength for each chunk is $\Theta(\sqrt{n} \log n)$. First, we encode using an RS outer code to create “coded-chunks” from the message chunks. Next, we encode each chunk by using an independently drawn an ensemble of low-weight random codes [16] that has the property that the expected codeword weight for each chunk is $\Theta(\log n)$.

With the above concatenated construction, the reliability analysis proceeds along familiar lines (with some parameter tweaks). Proving deniability, perhaps not surprisingly, turns out to be much more challenging. The first complication is imposed by the outer code – the ensemble of codes that our construction generates has linear dependencies between the chunks. This breaks the analysis from [16] that critically relies on each bit of the codewords being generated independently. It is conceivable that since the code is known to Willie, he may test for these dependencies and be able to come up with clever estimators of the transmission status. To overcome this problem, we use a systematic Reed-Solomon code. This decomposition of the chunks into systematic chunks and parity chunks is helpful in two ways. Firstly, this ensures that, at the very least, the systematic chunks are independently generated (since these correspond to independent message bits). Secondly, this also lets us show that, from Willie’s perspective the conditional distribution of transmissions in the parity chunks (of the Reed-Solomon outer code) is essentially statistically independent of Willie’s observations of transmissions in the systematic chunks, thus preventing him from gaining any advantage in estimating Alice’s transmission status by using the dependencies.

A second, and more technical, challenge is to prove that with high probability, the code for each chunk is deniable. In prior works such as [16], this is proved by first showing that under the ensemble average distribution, the codebook is deniable and then using a concentration argument over to show that with high probability over the codebook generation, the distribution imposed by the actual codebook is close to the ensemble average. Our concatenated code, however, only contains a polynomially small number of codewords in each chunk, since the chunk length scales as $\Theta(\sqrt{n} \log n)$. Especially when p_b approaches⁶ p_w , we need to provide a more sensitive analysis to ensure polynomially many plausible codewords for Willie in each chunk, but (*w.h.p.*) only one for Bob. Finally, we need to carefully combine proofs of deniability in each chunk to get deniability for the overall code.

By following this intuition, our work proves that one can communicate reliably and deniably with the best known throughput [16], while reducing the computational complexity to be polynomial with the blocklength n .

III. MODEL

All logarithms in this paper are binary.

Channel model: The channel between the transmitter Alice and the legitimate receiver Bob is a BSC(p_b), and the channel between Alice and the malicious eavesdropper Willie is a BSC(p_w), where $p_w > p_b$ (note that without this asymmetry, whenever Bob can decode reliably, so can Willie). Alice’s transmission status is denoted by $\mathbf{T} \in \{0, 1\}$ and the message is denoted by $\mathbf{M} \in \{0\} \cup \{1, 2, \dots, N\}$. When Alice communicates with Bob, her transmission status $\mathbf{T} = 1$ and the transmitted message \mathbf{M} is chosen uniformly at random from $\{1, 2, \dots, N\}$. When Alice does not communicate with Bob, her transmission status $\mathbf{T} = 0$ and the default message $\mathbf{M} = 0$ is transmitted. All three parties know the channel parameters p_b and p_w , but only Alice knows the transmission status \mathbf{T} and the message \mathbf{M} *a priori*. Figure 1 illustrates the system diagram of the communication model.

Encoder: Alice’s encoder is defined through the encoding function $\Psi(\cdot) : \{0\} \cup \{1, 2, \dots, N\} \rightarrow \{0, 1\}^n$, that is applied on the message \mathbf{M} to obtain the length- n binary codeword $\vec{\mathbf{X}} = \Psi(\mathbf{M})$. In particular, the innocent message $\mathbf{M} = 0$ will always be encoded to the length- n zero vector, *i.e.*, $\vec{\mathbf{X}} = \vec{\mathbf{0}}$. We define the rate of the code as $R = (\log N)/n$, and the relative throughput as $r = (\log N)/\sqrt{n}$. It is preferable to use the relative throughput r because when n goes to infinity, the relative throughput r scales as a constant while the rate R goes to zero.

Decoder: Bob receives the length- n binary vector $\vec{\mathbf{Y}}_b = \vec{\mathbf{X}} \oplus \vec{\mathbf{Z}}_b$, where $\vec{\mathbf{Z}}_b$ is the noise vector induced by the BSC(p_b), and applies a decoder map $\Gamma(\cdot) : \{0, 1\}^n \rightarrow \{0\} \cup \{1, 2, \dots, N\}$ to reconstruct the message $\hat{\mathbf{M}}$ from his observation $\vec{\mathbf{Y}}_b$. The goal is to guarantee the communication is $(1 - \epsilon_r)$ -reliable, *i.e.*, $\Pr_{\vec{\mathbf{Z}}_b}(\hat{\mathbf{M}} \neq \mathbf{M}) < \epsilon_r$.

Estimator: Willie aims to estimate \mathbf{T} from his observation $\vec{\mathbf{Y}}_w = \vec{\mathbf{X}} \oplus \vec{\mathbf{Z}}_w$, where $\vec{\mathbf{Z}}_w$ is the noise vector induced by the BSC(p_w), by using an estimator $\Phi(\cdot) : \{0, 1\}^n \rightarrow \{0, 1\}$ that outputs the estimate $\hat{\mathbf{T}} = \Phi(\vec{\mathbf{Y}}_w)$ of the transmission status. We use a *hypothesis-testing metric* to measure the deniability of the communication. Let $\alpha(\Phi) = \Pr_{\vec{\mathbf{Z}}_w}(\hat{\mathbf{T}} = 1 | \mathbf{T} = 0)$ be the

⁶As p_b approaches p_w , the chunk length grows accordingly.

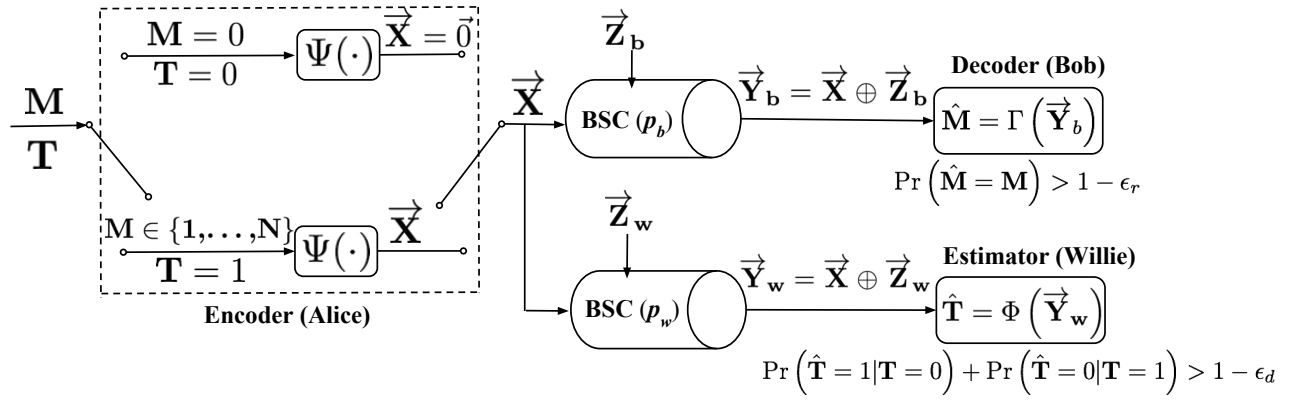


Fig. 1: *Reliable-Deniable Communication system diagram*: Alice has a message \mathbf{M} that can take N values $\{1, \dots, N\}$, and the transmission status $\mathbf{T} \in \{0, 1\}$. If Alice's transmission status $\mathbf{T} = 0$, she is required to “stay silent” – transmit the all zero codeword 0^n – this corresponds to the 0 message. On the other hand, if her transmission status $\mathbf{T} = 1$, she uses her encoder Ψ to encode her message \mathbf{M} into a codeword $\vec{\mathbf{X}}$. This $\vec{\mathbf{X}}$ is broadcast to the legitimate receiver Bob, and the eavesdropper Willie, over a pair of independent Binary Symmetric Channels with respective crossover probabilities p_b and p_w (respectively denoted by $\text{BSC}(p_b)$ and $\text{BSC}(p_w)$), which add Bernoulli noise vectors $\vec{\mathbf{Z}}_b$ and $\vec{\mathbf{Z}}_w$ respectively to $\vec{\mathbf{X}}$, resulting in the transmissions $\vec{\mathbf{Y}}_b$ and $\vec{\mathbf{Y}}_w$ observed respectively by Bob and Willie. Bob uses a decoder Γ to estimate Alice's transmitted message \mathbf{M} as $\hat{\mathbf{M}}$, and wishes to ensure reliability, i.e. that the probability (over channel noise $\vec{\mathbf{Z}}_b$) that $\hat{\mathbf{M}} \neq \mathbf{M}$ is “small” (at most ϵ_r). As a by-product of his decoder, he should therefore also detect Alice's transmission status. Willie, on the other hand, only wishes to decode Alice's transmission status \mathbf{T} . A code that is $(1 - \epsilon_d)$ -deniable ensures that, regardless of Willie's estimator, the probability (over Alice's message \mathbf{M} and channel noise $\vec{\mathbf{Z}}_b$) that $\Pr(\hat{\mathbf{T}} = 1 | \mathbf{T} = 0) + \Pr(\hat{\mathbf{T}} = 0 | \mathbf{T} = 1) > 1 - \epsilon_d$.

probability of *false alarm*, and $\beta(\Phi) = \Pr_{\mathbf{M}, \vec{\mathbf{Z}}_w}(\hat{\mathbf{T}} = 0 | \mathbf{T} = 1)$ be the probability of *missed detection*. The communication is deemed to be $(1 - \epsilon_d)$ -deniable if there does not exist an estimator Φ such that $\alpha(\Phi) + \beta(\Phi) < 1 - \epsilon_d$.

IV. MAIN RESULT

Before stating the main theorem, we first define an auxiliary function

$$f(x) = \log e - (1 + x) \log(e/(1 + x)). \quad (1)$$

Given any $0 < p_b < p_w < 1/2$ and sufficiently small $\epsilon_d > 0$, we define a *code weight design parameter*

$$k_2(p_w, \epsilon_d) = 2\epsilon_d \sqrt{p_w(1 - p_w)} / (1 - 2p_w), \quad (2)$$

and a *throughput parameter*

$$r_u = 2\epsilon_d \sqrt{p_w(1 - p_w)} \frac{1 - 2p_b}{1 - 2p_w} \log \left(\frac{1 - p_b}{p_b} \right). \quad (3)$$

The value of the code weight design parameter $k_2(p_w, \epsilon_d)$ is chosen to satisfy equations (26)-(28) in Section VIII, and the value of throughput parameter r_u is chosen to satisfy equations (104)-(105) in Section IX. Then we define four multivariable functions $g_i(u, v, w, t)$, where $1 \leq i \leq 4$, as

$$g_1(u, v, w, t) = k_2(u, v) \left[u(1 - w) \left(\log \left(\frac{1 - u}{u(1 - w)} \right) + \log e \right) + (1 - u)(1 + t) \left(\log \left(\frac{u}{(1 - u)(1 + t)} \right) + \log e \right) - \log e \right], \quad (4)$$

$$g_2(u, v, w, t) = k_2(u, v) \left[u(1 + w) \left(\log \left(\frac{1 - u}{u(1 + w)} \right) + \log e \right) + (1 - u)(1 + t) \left(\log \left(\frac{u}{(1 - u)(1 + t)} \right) + \log e \right) - \log e \right], \quad (5)$$

$$g_3(u, v, w, t) = k_2(u, v) \left[u(1 - w) \left(\log \left(\frac{1 - u}{u(1 - w)} \right) + \log e \right) + (1 - u)(1 - t) \left(\log \left(\frac{u}{(1 - u)(1 - t)} \right) + \log e \right) - \log e \right], \quad (6)$$

$$g_4(u, v, w, t) = k_2(u, v) \left[u(1 + w) \left(\log \left(\frac{1 - u}{u(1 + w)} \right) + \log e \right) + (1 - u)(1 - t) \left(\log \left(\frac{u}{(1 - u)(1 - t)} \right) + \log e \right) - \log e \right], \quad (7)$$

The reason why we define the multivariable functions $g_i(u, v, w, t)$ will be clear in equation (47), Section VIII. Equipped with the auxiliary tools above, we then define another *code chunk length design parameter* $k_1 = x(p_b, p_w, \epsilon_d)$ that is the smallest

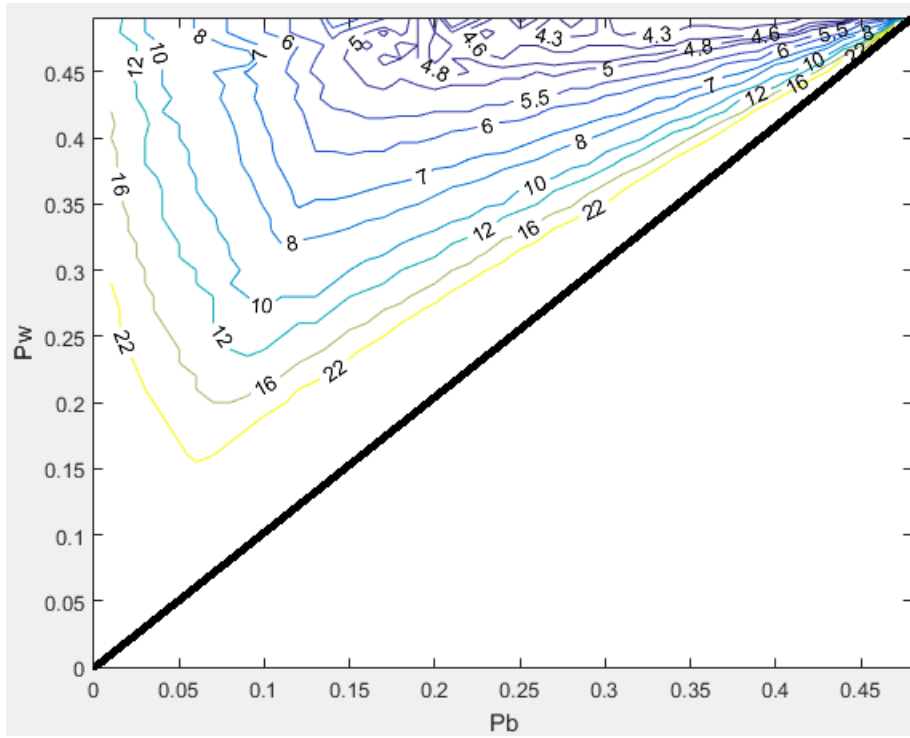


Fig. 2: This contour plot is obtained by finding the smallest non-zero value of the chunk-length parameter k_1 subject to constraints (8)-(11), and showing the decoding complexity of our concatenated code designs as a function of k_1 for various values of (p_b, p_w) . Each point on a contour labelled η corresponds to a (p_b, p_w) value with decoding complexity $\mathcal{O}(n^\eta)$. Our codes are only designed for the regime $p_b < p_w$ (less noisy channel to Bob than to Willie). As is to be expected, when p_b is close to p_w , the computational complexity is high (since the channels to both parties are similar, one has to employ longer block-lengths to be able to utilize the slight asymmetries in the two channels, leading to correspondingly higher computational cost). Interestingly, even in the regime when p_b is much smaller than p_w , the computational cost is also relatively high - in this regime the driving factor is the fact that a much higher deniable throughput is possible, leading to correspondingly higher computational workload.

positive value of x satisfying the following inequalities.

$$xr_u + x \max_{i=1}^4 \{g_i(p_w, \epsilon_d, \Delta_{10}^w, \Delta_{11}^w)\} \geq \frac{3}{2} + \delta \quad (8)$$

$$xp_w \cdot k_2(p_w, \epsilon_d) \cdot f(\Delta_{10}^w) \geq \frac{1}{2} + \delta \quad (9)$$

$$x(1 - p_w) \cdot k_2(p_w, \epsilon_d) \cdot f(\Delta_{11}^w) \geq \frac{1}{2} + \delta \quad (10)$$

$$0 < \Delta_{10}^w, \Delta_{11}^w < 1, \quad (11)$$

where δ is a *slackness parameter* that trades off the probability that a randomly chosen code is “good” with the computational complexity of encoding/decoding – this tradeoff can be seen in Figure BLAH. It can be chosen to be any value in the interval $(0, 1/2)$. For correctness we set $\delta = 0.01$ throughout this work. The parameters Δ_{10}^w and Δ_{11}^w , to be formally defined in Section VII-A, play an critical role in our code design. We elaborate on the reasons why Δ_{10}^w and Δ_{11}^w are required to satisfy inequalities (8)-(11) in equations (48), (61) and (62), Section VIII.

The work of [16] shows that given p_w, p_b and ϵ_d , one can transmit up to $r_u \sqrt{n}$ message bits per n channel uses deniably and reliably, but the decoding complexity as well as the space complexity for storing the codebook are exponential in \sqrt{n} . Our main result, Theorem 1 below, shows that it is possible to communicate reliably and deniably while reducing the complexity to be polynomial in n , by using a carefully designed concatenated code \mathcal{C}_n chosen from the concatenated code ensemble \mathcal{C}_n^{cc} (for notational convenience we drop the subscript n in the following) with throughput $r_u(1 - o(1))$.

Theorem 1. *Let k_1 be the smallest positive value satisfying Equations (8)-(11) and k_2 be as defined in Equation (2). For any $0 < p_b < p_w < 1/2$ and any sufficiently small $\epsilon_d > 0$, there exists a concatenated code ensemble \mathcal{C}^{cc} and a N_{p_b, p_w, ϵ_d} such that for any $n > N_{p_b, p_w, \epsilon_d}$, with probability at least $1 - \mathcal{O}(n^{-\delta}/\log n)$ over the concatenated code ensemble \mathcal{C}^{cc} , a randomly chosen code \mathcal{C} satisfies the following properties:*

- 1) *The relative throughput of the code r is $r_u \left(1 - \frac{1}{(\log n)^{1/3}}\right) = 2\epsilon_d \sqrt{p_w(1-p_w)} \frac{1-2p_b}{1-2p_w} \left(\log \frac{1-p_b}{p_b}\right) \left(1 - \frac{1}{(\log n)^{1/3}}\right)$.*
- 2) *There exists a decoder $\Gamma(\cdot)$ such that the reliability of the code is at least $1 - \exp(-2\sqrt{n}/(k_1(\log n)^2)) = 1 - \exp(-\mathcal{O}(\sqrt{n}/(\log n)^2))$.*
- 3) *The code is at least $(1 - \epsilon_d - 2n^{-\delta/4})$ -deniable from Willie.*

TABLE I: Effect of code design parameters on properties of the code

| Parameter | Code Property | Value |
|-----------|-----------------------------|------------------------|
| k_1 | Chunk length | $k_1 \sqrt{n} \log(n)$ |
| k_2 | Average weight of codewords | $k_2 \sqrt{n}$ |
| r_u | Throughput | $r_u \sqrt{n}$ |

- 4) The computational complexity of Alice's encoding is at most $rn/(k_1(\log n)^2)$, and that of Bob's decoding is at most n^{rk_1+1} . The space complexity for storing the codebook is n^{rk_1+1} .

Remarks.

a) The meaning of code parameters, as formalized in our proof, is summarized in Table I. The choice of these parameters leads to various tradeoffs in the complexity-throughput-deniability space.

- 1) The parameter k_1 characterizes the chunk length of our inner codes (which equal $k_1 \sqrt{n} \log(n)$) - the smaller the k_1 , the lower the complexity of the codes. However, making k_1 too small leads to problems proving deniability and/or complexity. Hence the problem of finding low-complexity codes is posed as a (non-convex) optimization problem.
- 2) The codewords in our codebook have average Hamming weight $k_2 \sqrt{n}$ - the specific choice of k_2 matches that in the (computationally inefficient) code design in [16].
- 3) Δ_{10}^w and Δ_{11}^w are, roughly speaking, parameters quantifying the type-classes of codeword-noise pairs likeliest to cause problems for our code design.
- 4) The throughput of our codes equals $r_u \sqrt{n}$, which also matches the throughput in [16].
- 5) The function $f(\cdot)$ helps analyze atypicality of codewords.
- 6) The functions $g_i(\cdot, \cdot, \cdot, \cdot)$ helps analyze the deniability of Alice's code. Together, the (non-convex) optimization problem finds the shortest k_1 (and hence the codes with the lowest complexity) that have a "good" deniable throughput.

b) The parameter k_1 affects the encoding and decoding complexity of the code. The encoding complexity is dominated by the complexity of Reed-Solomon encoding, and is a decreasing function of k_1 , while the decoding complexity is dominated by the random inner code and is an increasing function of k_1 . However, making k_1 too small leads to problems proving deniability and/or complexity.

c) The parameter k_2 determines the deniability of our code and is chosen to match the corresponding weight parameter in [16]. Likewise, the value r_u also matches the throughput parameter in [16].

d) For a given value of p_b and p_w , the choice of parameters that minimizes the overall decoding complexity is found by minimizing k_1 subject to inequalities (8)-(11). Even though the optimization is non-convex, finding the optimal k_1 is still tractable due to the small number of free variables ($\Delta_{10}^w, \Delta_{11}^w, k_1$) and monotonicity of k_1 . In Figure 2, we plot the optimal value of complexities for $0 < p_b < p_w < 1/2$.

e) For a specific choice of (p_b, p_w) , the decoding complexity is independent of the deniability parameter ϵ_d , while the relative throughput scales linearly with ϵ_d .

V. CODE DESIGN

In this section, we elaborate on the construction of our concatenated code. Our key technique is to use a "low-weight" random code to guarantee deniability. To reduce the computational cost, we divide the length- $(\Theta(\sqrt{n}))$ message into $\Theta(\sqrt{n}/\log n)$ chunks, with each chunk containing $\Theta(\log n)$ message bits, and apply random inner codes to each of the chunks. In addition, we use a Reed-Solomon code as an outer code to ensure the probability of error decays with the blocklength n .

A. Outer encoder and Inner encoders

Figure 3 illustrates the structure of the outer encoder and the inner encoders. For the outer RS code, we divide the length- $(r\sqrt{n})$ binary vector corresponding to the message \mathbf{M} into λL chunks $\mathbf{M}^{(1)}, \mathbf{M}^{(2)}, \dots, \mathbf{M}^{(\lambda L)}$, where $L = \sqrt{n}/(k_1 \log n)$ is the number of chunks, and λ is the rate of the outer code, with value specified below⁷. Therefore, each chunk contains $(rk_1/\lambda) \log n$ message bits. Let $r' = rk_1/\lambda$, and we regard each chunk as a symbol over finite field \mathbb{F}_q where $q = 2^{r' \log n}$. The encoding function of the outer code Ψ_{out} takes the form $\Psi_{out}(\cdot) : \mathbb{F}_q^{\lambda L} \rightarrow \mathbb{F}_q^L$, and we have $\Psi_{out}([\mathbf{M}^{(1)}, \mathbf{M}^{(2)}, \dots, \mathbf{M}^{(\lambda L)}]) = [\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \dots, \mathbf{W}^{(L)}]$. The first λL chunks $\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \dots, \mathbf{W}^{(\lambda L)}$ are *systematic chunks* while the last $(1 - \lambda)L$ chunks $\mathbf{W}^{(\lambda L+1)}, \dots, \mathbf{W}^{(L)}$ are *parity chunks*, since we use a systematic RS code as the outer code. Note that $\mathbf{W}^{(i)} = \mathbf{M}^{(i)}$ for the systematic chunks. In this work, we set the number of parity chunks to equal $56L/(\log n)$, and hence $\lambda = 1 - 56/(\log n)$ approaches 1 as n grows without bound.

⁷While a detailed discussion for this precise choice of the parameter λ is best left to Section IX, where the effect of the choice of the parameter is more apparent, for now it suffices to think of each systematic chunk as having a vanishing probability of decoding error, and hence a vanishingly small fraction of parity chunks sufficing to aid Bob's decoder.

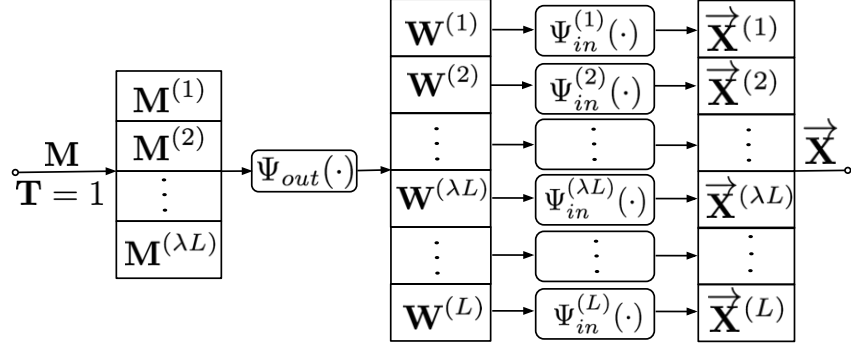


Fig. 3: The encoder of the concatenated code: Alice first divides the message M into λL chunks $M^{(1)}, M^{(2)}, \dots, M^{(\lambda L)}$. The outer encoder (corresponding to Reed-Solomon code) Ψ_{out} takes the λL chunks as input and outputs L chunks $W^{(1)}, W^{(2)}, \dots, W^{(L)}$, where $W^{(i)} = M^{(i)}$ for $1 \leq i \leq \lambda L$. For each chunk $W^{(i)}$, the inner encoder $\Psi_{in}^{(i)}$ (corresponding to the randomly-generated “low-weight” inner code $\mathcal{C}^{(i)}$) takes $W^{(i)}$ as input and outputs an inner codeword $\vec{X}^{(i)}$ of this inner code. The codeword \vec{X} of the concatenated code is obtained by collecting all the L inner codewords.

We use randomly generated “low-weight” inner codes to encode each of the chunks. The lengths of one inner codewords equal $n' = k_1 \sqrt{n} \log n$ since we have $L = \sqrt{n} / (k_1 \log n)$ chunks in total. For the i -th chunk, we generate a code $\mathcal{C}^{(i)}$ containing $2^{r' \log n}$ length- n' codewords, with each bit of these codewords chosen independently and identically distributed (*i.i.d.*) according to $\text{Bernoulli}(k_2 / \sqrt{n})$.

The probability distribution induced over concatenated codebooks generated via this process will be denoted $p(\mathcal{C}^{cc})$. Note that this differs from the probability distribution over *i.i.d.* codebook designs since, for example, the probability on any systematic inner codeword $\vec{X}^{(i)}$ is independent of any other systematic inner codeword $\vec{X}^{(j)}$, $j \neq i$. The encoder of $\mathcal{C}^{(i)}$ takes the form $\Psi_{in}^{(i)}(\cdot) : \{0, 1\}^{r' \log n} \rightarrow \{0, 1\}^{k_1 \sqrt{n} \log n}$, and outputs a length- n' vector $\vec{X}^{(i)} = \Psi_{in}^{(i)}(W^{(i)})$. The codebooks $\mathcal{C}^{(i)}$ are independently and identically distributed (*i.i.d.*) for all $i \in \{1, 2, \dots, L\}$ and hence different chunks are encoded by different inner codes. By collecting all the L inner codewords, we get $\vec{X} = [\vec{X}^{(1)}, \vec{X}^{(2)}, \dots, \vec{X}^{(L)}]$. In our concatenated code, $W^{(i)}$ is the output of the RS outer code, and also serves as the role of “message” of the inner random code. As a consequence, we name $W^{(i)}$ as *inner-message*.

B. Outer decoder and Inner decoder

Bob first partitions the received codeword \vec{Y}_b into $[\vec{Y}_b^{(1)}, \vec{Y}_b^{(2)}, \dots, \vec{Y}_b^{(L)}]$. The i -th inner decoder takes $\vec{Y}_b^{(i)}$ as input and reconstructs $\hat{W}^{(i)}$ by using the decoding function $\Gamma_{in}^{(i)}(\cdot) : \{0, 1\}^{k_1 \sqrt{n} \log n} \rightarrow \{0, 1\}^{r' \log n}$. Bob then treats each reconstructed length- $(r' \log n)$ vector $\hat{W}^{(i)}$ as a symbol over finite field \mathbb{F}_q , and then reconstructs \hat{M} using the decoder for a systematic RS code.

VI. PROBABILITY DISTRIBUTIONS OF INTEREST

As noted in the codebook design section, each inner code comprises of $n^{r'}$ codeword chunks, each of length n' . For each chunk i , the inner codebook $\mathcal{C}^{(i)}$ is chosen by choosing each bit of each of the codeword chunks *i.i.d.* according to $\text{Bernoulli}(k_2 / \sqrt{n})$. Hence the probability of a particular codebook $\mathcal{C}^{(i)}$, denoted $p(\mathcal{C}^{(i)})$, equals

$$\left(\frac{k_2}{\sqrt{n}} \right)^{wt_H(\mathcal{C}^{(i)})} \left(1 - \frac{k_2}{\sqrt{n}} \right)^{n^{r'} \cdot n' - wt_H(\mathcal{C}^{(i)})}. \quad (12)$$

The probability $p(\vec{x}^{(i)})$ that codeword chunk $\vec{x}^{(i)}$ is transmitted equals⁸

$$\frac{1}{2^{r' \log n}} = \frac{1}{n^{r'}}.$$

The probability $p(\vec{y}_w^{(i)} | \vec{x}^{(i)})$ that a transmitted codeword chunk $\vec{x}^{(i)}$ gets pushed by the $\text{Bernoulli}(p_w)$ noise on the channel to Willie to the channel output chunk $\vec{y}_w^{(i)}$, at Hamming distance $d_H(\vec{x}^{(i)}, \vec{y}_w^{(i)})$ from $\vec{x}^{(i)}$, thus equals

$$(p_w)^{d_H(\vec{x}^{(i)}, \vec{y}_w^{(i)})} (1 - p_w)^{n/L - d_H(\vec{x}^{(i)}, \vec{y}_w^{(i)})}.$$

⁸As is often the case in random code design, to make analysis cleaner, the summation is over the multiset corresponding to codeword chunks in $\mathcal{C}^{(i)}$ – due to random code design, with some (small) probability the same $x^{(i)}$ may correspond to different messages in that chunk.

Hence, if Alice is transmitting, the probability $p_1^{(i)}(\vec{y}_w^{(i)})$ of Willie observing channel output $\vec{y}_w^{(i)}$ on chunk i , equals

$$p_1^{(i)}(\vec{y}_w^{(i)}) = \sum_{\vec{x} \in \mathcal{C}^{(i)}} p(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}), \quad (13)$$

which can be further expanded as

$$\sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)}} (p_w)^{d_H(\vec{x}^{(i)}, \vec{y}_w^{(i)})} (1 - p_w)^{n/L - d_H(\vec{x}^{(i)}, \vec{y}_w^{(i)})} \frac{1}{2^{r' \log(n)}}.$$

The *ensemble average* distribution $\mathbb{E}_{\mathcal{C}^{(i)}}(p_1^{(i)}(\vec{y}_w^{(i)}))$ on Willie's channel outputs equals

$$\mathbb{E}_{\mathcal{C}^{(i)}}(p_1^{(i)}(\vec{y}_w^{(i)})) = \sum_{\mathcal{C}^{(i)}} p(\mathcal{C}^{(i)}) \sum_{\vec{x} \in \mathcal{C}^{(i)}} p(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}). \quad (14)$$

Using the definition of $p(\mathcal{C}^{(i)})$ in Equation (12) above, it can be seen that this corresponds to a Binomial($n, n(p_w * (k_2/\sqrt{n}))$) distribution, with

$$\mathbb{E}_{\mathcal{C}^{(i)}}(p_1^{(i)}(\vec{y}_w^{(i)})) = (p_w * (k_2/\sqrt{n}))^{w_H(\vec{y}_w^{(i)})} (1 - p_w * (k_2/\sqrt{n}))^{n/L - w_H(\vec{y}_w^{(i)})},$$

where recall that $a * b$ denotes the *binary convolution* operation $a(1 - b) + b(1 - a)$, and $w_H(\vec{y}_w^{(i)})$ denotes the Hamming weight of $\vec{y}_w^{(i)}$. The *no transmission* distribution $p_0(\vec{y}_w^{(i)})$ on Willie's channel outputs is, in contrast, a Binomial(n, np_w) distribution, with

$$p_0^{(i)}(\vec{y}_w^{(i)}) = (p_w)^{w_H(\vec{y}_w^{(i)})} (1 - p_w)^{n/L - w_H(\vec{y}_w^{(i)})}. \quad (15)$$

VII. NOTATION AND DEFINITIONS

A. Definitions for deniability

For any given binary vector \vec{x} , we denote the *fractional Hamming weight* of \vec{x} by $f_{1*} \triangleq wt_H(\vec{x})/\|\vec{x}\|_0$, where $\|\vec{x}\|_0$ is the ℓ_0 -norm of the vector \vec{x} . If, as will be the case in this work, each bit of each codeword is chosen to equal 1 with probability $\rho \triangleq k_2/\sqrt{n}$, then the expected value of f_{1*} equals ρ . Similarly, we denote the fractional Hamming weight of Willie's received vector \vec{y}_w by $f_{*1}^w \triangleq wt_H(\vec{y}_w)/\|\vec{y}_w\|_0$. The expected value of f_{*1}^w equals $\rho * p_w$, since f_{1*}^w and f_{0*}^w equal ρ and $(1 - \rho)$ respectively, and the channel between Alice and Willie is a BSC(p_w). Since much of the analysis in this work is based on a “chunk-wise” manner, we define the *n' -letter narrow typical set* of $\vec{Y}_w^{(i)}$ over chunks of length $n' = k_1 \sqrt{n} \log n$ (recall that $\vec{X}^{(i)}$ and $\vec{Y}_w^{(i)}$ represent the transmitted codeword and Willie's received vector of the i -th chunk respectively) when $\mathbf{T} = 1$ as⁹

$$\mathcal{A}_{n'}^1(\vec{Y}_w^{(i)}) \triangleq \left\{ \vec{y}_w^{(i)} : f_{*1}^w \in [\rho * p_w \cdot (1 \pm \Delta_{*1}^w)] \right\}, \quad (16)$$

where Δ_{*1}^w scales as $\mathcal{O}(1/\sqrt{n})$. By choosing Δ_{*1}^w carefully, we ensure that such a narrow typical set is a high probability set (as is usually the case in information-theoretic proofs), and is also as “narrow” as possible (includes as few type-classes as possible – this turns out to be important since extremal type-classes in the narrow typical set dominate the performance of our codes). It can be seen via standard arguments that if Δ_{*1}^w were to decay as $o(n^{-1/4})$, then the corresponding set $\mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})$ would have a vanishing probability mass – scaling Δ_{*1}^w as $O(n^{-1/4})$ results in the “narrowest” possible typical set. In this work, we choose Δ_{*1}^w to scale as $n^{-1/4 + \delta/2}$, where the slackness parameter δ (chosen in the range $(0, 1/2)$) allows one to show sufficiently tight concentration of probability.

For each pair of $(\vec{x}^{(i)}, \vec{y}_w^{(i)})$, the fraction of $(0, 0)$, $(0, 1)$, $(1, 0)$ and $(1, 1)$ pairs in $(\vec{x}^{(i)}, \vec{y}_w^{(i)})$ is denoted by f_{00}^w , f_{01}^w , f_{10}^w and f_{11}^w respectively. Note that $f_{1*} = f_{10}^w + f_{11}^w$ and $f_{*1}^w = f_{01}^w + f_{11}^w$ from the above definitions. We define the *n' -letter narrow conditionally typical set* of $\vec{X}^{(i)}$ given a particular $\vec{y}_w^{(i)}$ as

$$\mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_w^{(i)}) \triangleq \left\{ (\vec{x}^{(i)}, \vec{y}_w^{(i)}) : \begin{array}{l} f_{10}^w \in [\rho p_w (1 \pm \Delta_{10}^w)] \\ f_{11}^w \in [\rho (1 - p_w) (1 \pm \Delta_{11}^w)] \end{array} \right\}, \quad (17)$$

where Δ_{10}^w and Δ_{11}^w scale as constants in the interval $[0, 1]$ (with values to be specified later, in Section VIII – indeed, careful choice of these two parameters turns out to be critical for our code design). The *n' -letter conditionally typical set* can further be decomposed to many *n' -letter conditional type classes*. The *n' -letter conditional type class* of $\vec{X}^{(i)}$ given $\vec{y}_w^{(i)}$ is defined as

$$\mathcal{T}_{n'}^1(\vec{X}^{(i)} | \vec{y}_w^{(i)})(f_{10}^w, f_{11}^w) \triangleq \left\{ (\vec{x}^{(i)}, \vec{y}_w^{(i)}) : \begin{array}{l} |j : (\vec{x}_j^{(i)}, \vec{y}_{w,j}^{(i)}) = (1, 0)| = n' f_{10}^w \\ |j : (\vec{x}_j^{(i)}, \vec{y}_{w,j}^{(i)}) = (1, 1)| = n' f_{11}^w \end{array} \right\}, \quad (18)$$

⁹For notational convenience, we use $[a \pm b]$ to denote an interval $[a - b, a + b] \subset \mathbf{R}$.

where $\vec{x}_j^{(i)}$ and $\vec{y}_{w,j}^{(i)}$ are the j -th elements of $\vec{x}^{(i)}$ and $\vec{y}_w^{(i)}$ respectively. Therefore, we can represent the n' -letter narrow conditionally typical set as the union of many “typical” conditional type classes, *i.e.*,

$$\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)}|\vec{y}_w^{(i)}) = \bigcup_{f_{10}^w \in [\rho p_w(1 \pm \Delta_{10}^w)] \text{ and } f_{11}^w \in [\rho(1-p_w)(1 \pm \Delta_{11}^w)]} \mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)}|\vec{y}_w^{(i)})(f_{10}^w, f_{11}^w).$$

B. Definitions for reliability

For the i -th chunk, we denote the fractional Hamming weight of Bob’s received vector $\vec{y}_b^{(i)}$ by $f_{*1}^b \triangleq \text{wt}_H(\vec{y}_b^{(i)})/n'$. The expected value of f_{*1}^b equals p_b when $\mathbf{T} = 0$, and equals $\rho * p_b$ when $\mathbf{T} = 1$. Then we define the n' -letter narrow typical set of $\vec{\mathbf{Y}}_b^{(i)}$ when $\mathbf{T} = 0$ as

$$\mathcal{A}_{n'}^0(\vec{\mathbf{Y}}_b^{(i)}) \triangleq \left\{ \vec{y}_b^{(i)} : f_{*1}^b \in \left[p_b(1 \pm \Delta_{*1}^{b,(0)}) \right] \right\}, \quad (19)$$

and the n' -letter narrow typical set of $\vec{\mathbf{Y}}_b^{(i)}$ when $\mathbf{T} = 1$ as

$$\mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) \triangleq \left\{ \vec{y}_b^{(i)} : f_{*1}^b \in \left[\rho * p_b \cdot (1 \pm \Delta_{*1}^{b,(1)}) \right] \right\}, \quad (20)$$

where $\Delta_{*1}^{b,(0)}$ and $\Delta_{*1}^{b,(1)}$ both scale as $\mathcal{O}(n^{-1/4+\delta/2})$ in the following proof¹⁰. Moreover, we also denote the fraction of $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$ pairs in $(\vec{x}^{(i)}, \vec{y}_b^{(i)})$ by f_{00}^b , f_{01}^b , f_{10}^b and f_{11}^b respectively. The n' -letter narrow conditionally typical set of $\vec{\mathbf{X}}^{(i)}$ given $\vec{y}_b^{(i)}$ when $\mathbf{T} = 1$ is defined as

$$\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)}|\vec{y}_b^{(i)}) \triangleq \left\{ (\vec{x}^{(i)}, \vec{y}_b^{(i)}) : \begin{array}{l} f_{10}^b \in [\rho p_b(1 \pm \Delta_{10}^b)] \\ f_{11}^b \in [\rho(1-p_b)(1 \pm \Delta_{11}^b)] \end{array} \right\}, \quad (21)$$

where Δ_{10}^b and Δ_{11}^b scale as $\mathcal{O}(1/(\log n)^{1/2})$. The scalings of Δ_{10}^b and Δ_{11}^b , which are analyzed in Claim 14 and Appendix B, guarantee simultaneously that the conditionally typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)}|\vec{y}_b^{(i)})$ is a high probability set, and yet is also as “narrow” as possible. We then define the n' -letter conditional type class of $\vec{\mathbf{X}}^{(i)}$ given $\vec{y}_b^{(i)}$ as

$$\mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)}|\vec{y}_b^{(i)})(f_{10}^b, f_{11}^b) \triangleq \left\{ (\vec{x}^{(i)}, \vec{y}_b^{(i)}) : \begin{array}{l} |j : (\vec{x}_j^{(i)}, \vec{y}_{b,j}^{(i)}) = (1, 0)| = n' f_{10}^b \\ |j : (\vec{x}_j^{(i)}, \vec{y}_{b,j}^{(i)}) = (1, 1)| = n' f_{11}^b \end{array} \right\}, \quad (22)$$

and therefore,

$$\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)}|\vec{y}_b^{(i)}) = \bigcup_{f_{10}^b \in [\rho p_b(1 \pm \Delta_{10}^b)] \text{ and } f_{11}^b \in [\rho(1-p_b)(1 \pm \Delta_{11}^b)]} \mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)}|\vec{y}_b^{(i)})(f_{10}^b, f_{11}^b).$$

For convenience, we refer to $\mathcal{A}_{n'}^1$ and $\mathcal{T}_{n'}^1$ as conditionally typical set and conditional type class respectively, if the length of vectors we are interested in is clear from the context.

C. Empirical mutual information and empirical KL divergence

For the i -th chunk, given the actual inner codeword $\vec{x}^{(i)}$ (each bit of $\vec{x}^{(i)}$ is generated *i.i.d.* according to Bernoulli(ρ)), the empirical Kullback-Leibler divergence between the actual inner codeword $\vec{x}^{(i)}$ and the code design parameter ρ is defined as

$$D(\vec{x}^{(i)} \parallel \rho) \triangleq f_{0*} \log \frac{f_{0*}}{1-\rho} + f_{1*} \log \frac{f_{1*}}{\rho}. \quad (23)$$

Given the actual inner codeword $\vec{x}^{(i)}$, Willie’s received vector $\vec{y}_w^{(i)}$ and Bob’s received vector $\vec{y}_b^{(i)}$, the empirical mutual information between $\vec{x}^{(i)}$ and $\vec{y}_w^{(i)}$ is defined as

$$I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) \triangleq \sum_{(j,j') \in \{0,1\} \times \{0,1\}} f_{jj'}^w \log \frac{f_{jj'}^w}{f_{j*}^w \cdot f_{*j'}^w}, \quad (24)$$

and the empirical mutual information between $\vec{x}^{(i)}$ and $\vec{y}_b^{(i)}$ is defined as

$$I(\vec{x}^{(i)}; \vec{y}_b^{(i)}) \triangleq \sum_{(j,j') \in \{0,1\} \times \{0,1\}} f_{jj'}^b \log \frac{f_{jj'}^b}{f_{j*}^b \cdot f_{*j'}^b}. \quad (25)$$

The empirical mutual information $I(\vec{x}^{(i)}; \vec{y}_w^{(i)})$ is uniquely determined by the triplet $(f_{*1}^w, f_{10}^w, f_{11}^w)$, since f_{00}^w and f_{01}^w can be computed from this triplet. The same argument works for $I(\vec{x}^{(i)}; \vec{y}_b^{(i)})$. The empirical Kullback-Leibler divergence between $\vec{x}^{(i)}$ and ρ is uniquely determined by the pair (f_{10}^w, f_{11}^w) since it is only related to the vector $\vec{x}^{(i)}$.

¹⁰The reason for this scaling is as in Section VIII-A for Δ_{*1}^w .

TABLE II: Table of Parameters

| Symbol | Description | Equality/Range | Section |
|--|---|--|--------------------------|
| \mathbf{M} | message | $\mathbf{M} \in \{1, \dots, N\}$ | Section III |
| \mathbf{T} | transmission status | $\mathbf{T} \in \{0, 1\}$ | Section III |
| $\vec{\mathbf{X}}$ | codeword | $\vec{\mathbf{X}} \in \{0, 1\}^n$ | Section III |
| $\vec{\mathbf{Y}}_b(\vec{\mathbf{Y}}_w)$ | received vector of Bob (Willie) | $\vec{\mathbf{Y}}_b \in \{0, 1\}^n$ ($\vec{\mathbf{Y}}_w \in \{0, 1\}^n$) | Section III |
| $\vec{\mathbf{Z}}_b(\vec{\mathbf{Z}}_w)$ | error vector of Bob (Willie) | $\vec{\mathbf{Z}}_b \in \{0, 1\}^n$ ($\vec{\mathbf{Z}}_w \in \{0, 1\}^n$) | Section III |
| p_b | crossover probability of BSC between Alice and Bob | $p_b \in [0, 0.5]$ | Section I |
| p_w | crossover probability of BSC between Alice and Willie | $p_w \in [0, 0.5]$ | Section I |
| $\Psi(\cdot)$ | encoder | $\Psi(\cdot) : \{0\} \cup \{1, 2, \dots, N\} \rightarrow \{0, 1\}^n$ | Section III |
| $\Gamma(\cdot)$ | Bob's decoder | $\Gamma(\cdot) : \{0, 1\}^n \rightarrow \{0\} \cup \{1, 2, \dots, N\}$ | Section III |
| $\Phi(\cdot)$ | Willie's estimator | $\Phi(\cdot) : \{0, 1\}^n \rightarrow \{0, 1\}$ | Section III |
| $\alpha(\Phi)$ | probability of false alarm | $\alpha(\Phi) = \Pr_{\vec{\mathbf{Z}}_w}(\hat{\mathbf{T}} = 1 \mathbf{T} = 0)$ | Section III |
| $\beta(\Phi)$ | probability of missed detection | $\beta(\Phi) = \Pr_{\mathbf{M}, \vec{\mathbf{Z}}_w}(\hat{\mathbf{T}} = 0 \mathbf{T} = 1)$ | Section III |
| \mathcal{C} | concatenated code | | Section V |
| \mathcal{C}^{cc} | concatenated code ensemble | | Section V |
| $p(\mathcal{C}^{cc})$ | probability distribution of concatenated codes | | Section V-A |
| ϵ_d | parameter of deniability | $0 < \epsilon_d < 1$ | Section III |
| ϵ_r | parameter of reliability | $0 < \epsilon_r < 1$ | Section III |
| R | rate of the concatenated code | $R = (\log N)/n$ (goes to 0) | Section III |
| r | relative throughput of the concatenated code (achieved) | $r = (\log N)/\sqrt{n}$ | Section III |
| k_1 | parameter of chunk length | smallest positive value satisfying (5)-(8) | Section IV |
| n' | chunk length | $n' = k_1 \sqrt{n \log n}$ | Section V-A |
| k_2 | average weight of codewords | $k_2 = 2\epsilon_d \sqrt{p_w(1-p_w)/(1-2p_w)}$ | Equation (2), Section IV |
| ρ | average fraction of 1's in codewords | $\rho = k_2/\sqrt{n}$ | Section VII-A |
| r_u | maximum relative throughput | $r_u = 2\epsilon_d \sqrt{p_w(1-p_w)} \frac{1-2p_b}{1-2p_w} \log\left(\frac{1-p_b}{p_b}\right)$ | Equation (3), Section IV |
| L | number of chunks | $L = \sqrt{n}/(k_1 \log n)$ | Section V-A |
| λ | rate of the outer code | $\lambda = 1 - o(1)$ | Section V-A |
| l_1 | number of systematic chunks | $l_1 = L\lambda$ | Section VIII-A |
| l_2 | number of parity chunks | $l_2 = L(1 - \lambda)$ | Section VIII-A |
| r' | relative throughput of an inner code | $r' = rk_1/\lambda$ | Section V-A |
| \mathbb{F}_q | finite field of the outer RS code | $q = 2^{r' \log n}$ | Section V-A |
| $\Psi_{out}(\cdot)$ | outer encoder | $\Psi_{out}(\cdot) : \mathbb{F}_q^{\lambda L} \rightarrow \mathbb{F}_q^L$ | Section V-A |
| $\Psi_{in}^{(i)}(\cdot)$ | inner encoder of the i -th chunk | $\Psi_{in}^{(i)}(\cdot) : \{0, 1\}^{r' \log n} \rightarrow \{0, 1\}^{n'}$ | Section V-A |
| $\mathcal{C}^{(i)}$ | inner code of the i -th chunk | | Section V-A |
| $\mathbf{W}^{(i)}$ | inner-message of the i -th chunk | $\mathbf{W}^{(i)} \in \{0, 1\}^{r' \log n}$ | Section V-A |
| $\vec{\mathbf{X}}^{(i)}$ | codeword of the i -th chunk | $\vec{\mathbf{X}}^{(i)} \in \{0, 1\}^{n'}$ | Section V-A |
| f_{ij}^w | fraction of pair- (i, j) in (\vec{x}, \vec{y}_w) | $0 \leq f_{ij}^w \leq 1, i, j \in \{0, 1\}$ | Section VII-A |
| $\mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})$ | n' -letter narrow typical set of $\vec{\mathbf{Y}}_w^{(i)}$ ($\mathbf{T} = 1$) | | Term (16), Section VII-A |
| $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} \vec{y}_w^{(i)})$ | n' -letter conditionally typical set of $\vec{\mathbf{X}}^{(i)}$ given $\vec{y}_w^{(i)}$ ($\mathbf{T} = 1$) | | Term (17), Section VII-A |
| $\mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)} \vec{y}_w^{(i)})$ | n' -letter conditional type class of $\vec{\mathbf{X}}^{(i)}$ given $\vec{y}_w^{(i)}$ ($\mathbf{T} = 1$) | | Term (18), Section VII-A |
| f_{ij}^b | fraction of pair- (i, j) in (\vec{x}, \vec{y}_b) | $0 \leq f_{ij}^b \leq 1, i, j \in \{0, 1\}$ | Section VII-B |
| $\mathcal{A}_{n'}^0(\vec{\mathbf{Y}}_b^{(i)})$ | n' -letter narrow typical set of $\vec{\mathbf{Y}}_b^{(i)}$ ($\mathbf{T} = 0$) | | Term (19), Section VII-B |
| $\mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)})$ | n' -letter narrow typical set of $\vec{\mathbf{Y}}_b^{(i)}$ ($\mathbf{T} = 1$) | | Term (20), Section VII-B |
| $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} \vec{y}_b^{(i)})$ | n' -letter conditionally typical set of $\vec{\mathbf{X}}^{(i)}$ given $\vec{y}_b^{(i)}$ ($\mathbf{T} = 1$) | | Term (21), Section VII-B |
| $\mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)} \vec{y}_b^{(i)})$ | n' -letter conditional type class of $\vec{\mathbf{X}}^{(i)}$ given $\vec{y}_b^{(i)}$ ($\mathbf{T} = 1$) | | Term (22), Section VII-B |
| $D(\vec{x}^{(i)} \ \rho)$ | empirical Kullback-Leibler divergence between $\vec{x}^{(i)}$ and ρ | $D(\vec{x} \ \rho) \triangleq f_{0*} \log \frac{f_{0*}}{1-\rho} + f_{1*} \log \frac{f_{1*}}{f_{jj'}^w \rho}$ | Term (23), Section VII-C |
| $I(\vec{x}^{(i)}; \vec{y}_w^{(i)})$ | empirical mutual information between $\vec{x}^{(i)}$ and $\vec{y}_w^{(i)}$ | $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) \triangleq \sum_{(j,j')} f_{jj'}^w \log \frac{f_{jj'}^w}{f_{j*} \cdot f_{*j'}}$ | Term (24), Section VII-C |
| $I(\vec{x}^{(i)}; \vec{y}_b^{(i)})$ | empirical mutual information between $\vec{x}^{(i)}$ and $\vec{y}_b^{(i)}$ | $I(\vec{x}^{(i)}; \vec{y}_b^{(i)}) \triangleq \sum_{(j,j')} f_{jj'}^b \log \frac{f_{jj'}^b}{f_{j*} \cdot f_{*j'}}$ | Term (25), Section VII-C |
| p_0 | innocent distribution of \vec{y}_w when $\mathbf{T} = 0$ | | Term (15), Section VI |
| p_1 | active distribution of \vec{y}_w when $\mathbf{T} = 1$ | | Term (13), Section VI |
| $\mathbb{E}_{\mathcal{C}}(p_1)$ | "ensemble-averaged" active distribution of \vec{y}_w when $\mathbf{T} = 1$ | $\mathbb{E}_{\mathcal{C}}(p_1)(\vec{y}_w) = \mathbb{E}_{\mathcal{C}}(p_1(\vec{y}_w))$ | Term (14), Section VI |

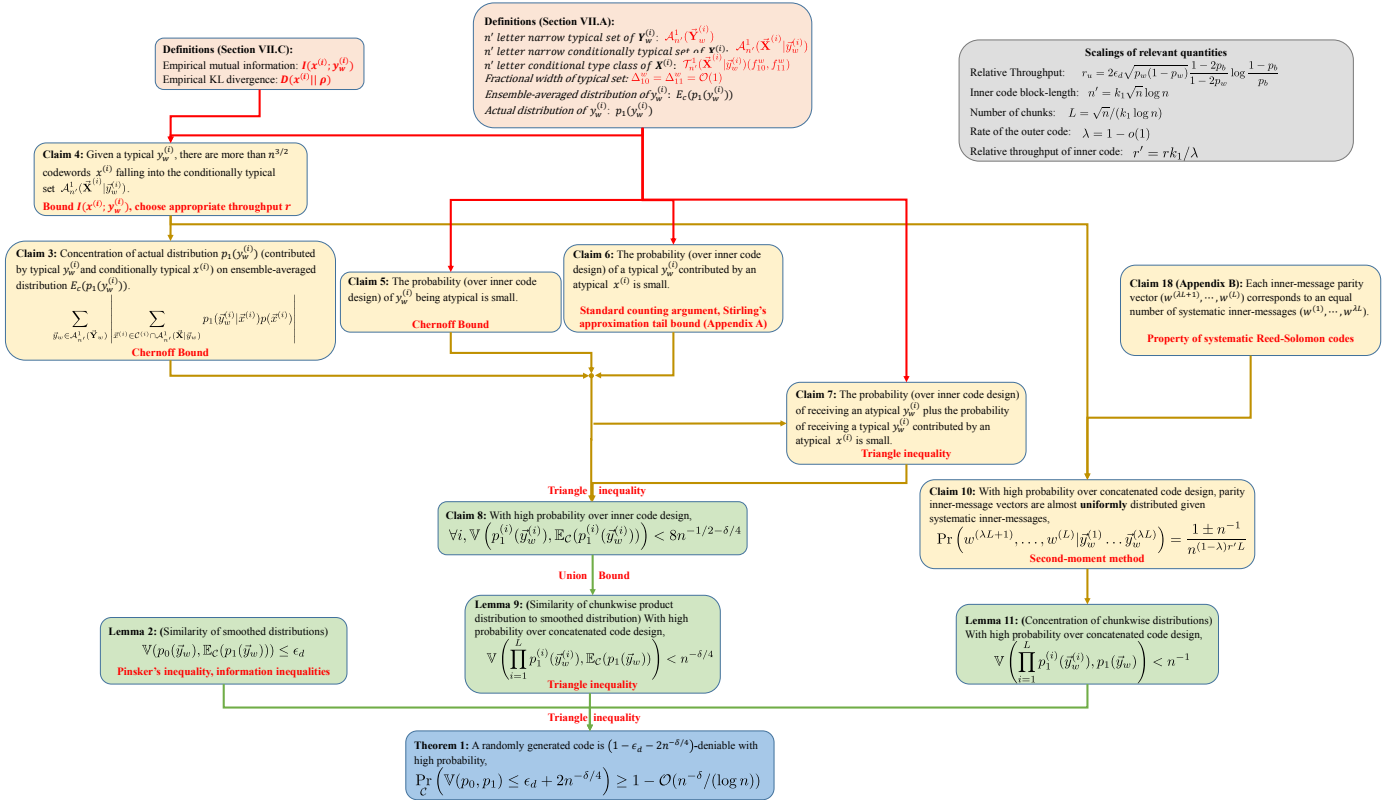


Fig. 4: A road-map of our proof that our codes are highly deniable with high probability.

VIII. PROOF OF DENIABILITY

Recall that the code is $(1 - \epsilon_d - 2n^{-\delta/4})$ -deniable if there does not exist an estimator Φ such that $\alpha(\Phi) + \beta(\Phi) < 1 - \epsilon_d - 2n^{-\delta/4}$. By “standard statistical arguments” [34], it is equivalent to say that $\mathbb{V}(p_0, p_1) \leq \epsilon_d + 2n^{-\delta/4}$, where p_0 stands for the *innocent distribution* of \mathbf{y}_w when Alice’s transmission status $\mathbf{T} = 0$ and p_1 stands for the *active distribution* of \mathbf{y}_w when Alice’s transmission status $\mathbf{T} = 1$. Furthermore, we define the “ensemble-averaged” active distribution of \mathbf{y}_w when $\mathbf{T} = 1$ as $E_c(p_1)$, where

$$E_c(p_1)(\mathbf{y}_w) = E_c(p_1(\mathbf{y}_w)) = \sum_{\mathcal{C}} \Pr(\mathcal{C}) \sum_{\mathbf{x} \in \mathcal{C}} p(\mathbf{y}_w | \mathbf{x}) p(\mathbf{x})$$

for all \mathbf{y}_w . Note that this can be viewed as passing the all-zero codeword through two successive BSCs, with crossover probabilities respectively ρ and p_w , and hence $E_c(p_1(\mathbf{y}_w))$ itself has a relatively simple description, corresponding to $(\rho * p_w)^{wt_H(\mathbf{y}_w)} (1 - \rho * p_w)^{n - wt_H(\mathbf{y}_w)}$, even though for specific codes $p_1(\mathbf{y}_w)$ has a complicated dependence on \mathcal{C} . Since variational distance satisfies the triangle inequality, we have

$$\mathbb{V}(p_0, p_1) \leq \mathbb{V}(p_0, E_c(p_1)) + \mathbb{V}(E_c(p_1), p_1).$$

Following the approach in [16], to prove that the proposed code is deniable, it suffices to show that (i) $\mathbb{V}(p_0, E_c(p_1)) < \epsilon_d$ and (ii) with sufficiently high probability over the design of the concatenated code’s inner codes, $\mathbb{V}(E_c(p_1), p_1) < 2n^{-\delta/4}$. A flow-chart of the proof of deniability can be found in Figure 4. As in [16], the proof of (i) follows fairly directly from relatively standard information-theoretic inequalities. For completeness, we repeat the proof here.

Lemma 2. [16] If the code design parameter $k_2 < \frac{2\epsilon_d \sqrt{p_w(1-p_w)}}{1-2p_w}$, then $\mathbb{V}(p_0, E_c(p_1)) \leq \epsilon_d$.

Proof:

$$\mathbb{V}(p_0, \mathbb{E}_{\mathcal{C}}(p_1)) \leq \sqrt{\frac{\ln 2}{2} D(p_0 \parallel \mathbb{E}_{\mathcal{C}}(p_1))} \quad (26)$$

$$= \sqrt{\frac{n \ln 2}{2} D(p_w \parallel \rho * p_w)} \quad (27)$$

$$\leq \sqrt{\frac{n \ln 2}{2} \left(\frac{\rho^2 (1 - 2p_w)^2}{2p_w (1 - p_w) \ln 2} + \mathcal{O}(\rho^3) \right)}, \quad (28)$$

where (26) follows from Pinsker's inequality and (27) follows from the chain rule for relative entropy, since both p_0 and $\mathbb{E}_{\mathcal{C}}(p_1)$ correspond to n -letter sequences drawn i.i.d. from $\text{Bernoulli}(p_w)$ and $\text{Bernoulli}(\rho * p_w)$ distributions respectively. The probability of a single bit of the distribution p_0 and $\mathbb{E}_{\mathcal{C}}(p_1)$ being 1 equals p_w and $\rho * p_w = \rho(1 - p_w) + (1 - \rho)p_w$ respectively. Equation (28) follows by taking the Taylor series expansion for KL-divergence, as in [16, Claim 3], resulting in $D(p_w \parallel \rho * p_w) \leq \frac{\rho^2 (1 - 2p_w)^2}{2p_w (1 - p_w) \ln 2} + \mathcal{O}(\rho^3)$. By choosing $\rho = \frac{k_2}{\sqrt{n}} < \frac{2\epsilon_d \sqrt{p_w(1-p_w)}}{(1-2p_w)\sqrt{n}}$, we have $\mathbb{V}(p_0, \mathbb{E}_{\mathcal{C}}(p_1)) < \epsilon_d + \mathcal{O}(n^{-1/4})$. \square

We now proceed to one of the major parts of our proof (proof of (ii)) – showing that with high probability over the choice of the inner codes, the variational distance between p_1 (which depends on the specific inner codes chosen) and the ensemble-averaged distribution $\mathbb{E}_{\mathcal{C}}(p_1)$ is small. As mentioned in the introduction, this is considerably more challenging in our setting of concatenated codes comprising of multiple chunks, than in the setting of [16] and other works, wherein a single n -letter code is used.

Recall that as described in Section V-A, the concatenated encoder generates a length- n codeword by generating $L = \sqrt{n}/(k_1 \log n)$ sub-codewords, each of length $n' = k_1 \sqrt{n} \log n$. Correspondingly, we partition the received vector \vec{y}_w into L length- n' vectors $\vec{y}_w^{(1)}, \vec{y}_w^{(2)}, \dots, \vec{y}_w^{(L)}$, where for each i the vector $\vec{y}_w^{(i)}$ corresponds to the set of channel outputs $y_{(i-1)n'+1}, \dots, y_{in'}$. For each chunk $i \in \{1, \dots, L\}$, we denote the *active* and the “ensemble-averaged” *active* n' -letter distribution of $\vec{y}_w^{(i)}$ when $\mathbf{T} = 1$ by $p_1^{(i)}$ and $\mathbb{E}_{\mathcal{C}}(p_1^{(i)})$ respectively, where

$$p_1^{(i)}(\vec{y}_w^{(i)}) = \sum_{\vec{x} \in \mathcal{C}^{(i)}} p(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}),$$

$$\mathbb{E}_{\mathcal{C}}(p_1^{(i)}(\vec{y}_w^{(i)})) = \sum_{\mathcal{C}^{(i)}} \Pr(\mathcal{C}^{(i)}) \sum_{\vec{x} \in \mathcal{C}^{(i)}} p(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}).$$

By the definition of variational distance, we have

$$\begin{aligned} & \mathbb{V}(\mathbb{E}_{\mathcal{C}}(p_1), p_1) \\ &= \frac{1}{2} \sum_{\vec{y}_w \in \{0,1\}^n} |\mathbb{E}_{\mathcal{C}}(p_1(\vec{y}_w)) - p_1(\vec{y}_w)| \\ &= \frac{1}{2} \sum_{\vec{y}_w^{(1)} \dots \vec{y}_w^{(L)}} \left| \mathbb{E}_{\mathcal{C}}(p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(L)})) - p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(L)}) \right| \\ &\leq \frac{1}{2} \sum_{\vec{y}_w^{(1)} \dots \vec{y}_w^{(L)}} \left| p_1^{(1)}(\vec{y}_w^{(1)}) \dots p_1^{(L)}(\vec{y}_w^{(L)}) - p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(L)}) \right| \\ &\quad + \frac{1}{2} \sum_{\vec{y}_w^{(1)} \dots \vec{y}_w^{(L)}} \left| \mathbb{E}_{\mathcal{C}}(p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(L)})) - p_1^{(1)}(\vec{y}_w^{(1)}) \dots p_1^{(L)}(\vec{y}_w^{(L)}) \right|. \end{aligned} \quad (29)$$

In Equation (29) above, the first term corresponds to the variational distance between the n -letter distribution p_1 on \vec{y}_w , and a corresponding “chunk-wise independent” product distribution denoted by $p_1^{(1)} p_1^{(2)} \dots p_1^{(L)}$; and the second term corresponds to the variational distance between the n -letter ensemble average distribution $\mathbb{E}_{\mathcal{C}}(p_1)$ on \vec{y}_w , and the same product distribution (the inequality in the last term follows from the triangle inequality). This latter distribution corresponds to the distribution that Willie would see if he were to “assume” that the distribution on \vec{y}_w splits as a product of independent distributions on $\vec{y}_w^{(1)}, \vec{y}_w^{(2)}, \dots, \vec{y}_w^{(L)}$. There is of course no reason for this to be the case, especially since Alice is using a code that introduces correlations between chunks, but introducing such a “proxy” distribution and computing variational distributions with respect to it is a useful analytical tool. Intuitively, for a highly deniable concatenated code, the product distribution $p_1^{(1)} p_1^{(2)} \dots p_1^{(L)}$ should be “close” to *both* the actual distribution p_1 , and the ensemble average distribution $\mathbb{E}_{\mathcal{C}}(p_1)$. Indeed, this is what we show below. We prove Lemma 3 and Lemma 4 in Section VIII-A and Section VIII-B respectively.

Lemma 3. With probability at least $1 - \sqrt{n} \exp(-\frac{4}{3}n^{1/2+\delta/2})$ over channel noise to Willie and concatenated code design, the ensemble-averaged distribution is close to the “chunk-wise independent” product distribution, i.e.,

$$\frac{1}{2} \sum_{\vec{y}_w^{(1)} \dots \vec{y}_w^{(L)}} \left| \mathbb{E}_{\mathcal{C}}(p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(L)})) - p_1^{(1)}(\vec{y}_w^{(1)}) \dots p_1^{(L)}(\vec{y}_w^{(L)}) \right| < n^{-\delta/4}.$$

Lemma 4. With probability at least $1 - \mathcal{O}(n^{-\delta/\log n})$ over channel noise to Willie and concatenated code design, the n -letter distribution p_1 on \vec{y}_w is close to the “chunk-wise independent” product distribution, i.e.,

$$\frac{1}{2} \sum_{\vec{y}_w^{(1)} \dots \vec{y}_w^{(L)}} \left| p_1^{(1)}(\vec{y}_w^{(1)}) \dots p_1^{(L)}(\vec{y}_w^{(L)}) - p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(L)}) \right| \leq \frac{1}{2} n^{-1}.$$

A. Proof of Lemma 3:

We first observe that the variational distance between the ensemble-averaged distribution and the “chunk-wise independent” product distribution is bounded from above as

$$\begin{aligned} & \frac{1}{2} \sum_{\vec{y}_w^{(1)} \dots \vec{y}_w^{(L)}} \left| \mathbb{E}_{\mathcal{C}}(p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(L)})) - p_1^{(1)}(\vec{y}_w^{(1)}) \dots p_1^{(L)}(\vec{y}_w^{(L)}) \right| \\ &= \frac{1}{2} \sum_{\vec{y}_w^{(1)} \dots \vec{y}_w^{(L)}} \left| \mathbb{E}_{\mathcal{C}^{(1)}}(p_1^{(1)}(\vec{y}_w^{(1)})) \dots \mathbb{E}_{\mathcal{C}^{(L)}}(p_1^{(L)}(\vec{y}_w^{(L)})) - p_1^{(1)}(\vec{y}_w^{(1)}) \dots p_1^{(L)}(\vec{y}_w^{(L)}) \right| \end{aligned} \quad (30)$$

$$\leq \frac{1}{2} \sum_{i=1}^L \sum_{\vec{y}_w^{(i)} \in \{0,1\}^{n'}} \left| \mathbb{E}_{\mathcal{C}^{(i)}}(p_1^{(i)}(\vec{y}_w^{(i)})) - p_1^{(i)}(\vec{y}_w^{(i)}) \right|. \quad (31)$$

Equation (30) follows since the inner codebooks for each chunk are generated i.i.d. and the channel is memoryless. Equation (31) follows from the triangle inequality. We now follow the lead of the analysis in [16] by replicating the analysis there in a chunk-wise manner. Specifically, for each chunk $i \in \{1, \dots, L\}$, we break up $\frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \{0,1\}^{n'}} \left| \mathbb{E}_{\mathcal{C}^{(i)}}(p_1^{(i)}(\vec{y}_w^{(i)})) - p_1^{(i)}(\vec{y}_w^{(i)}) \right|$, the variational distance between $p_1^{(i)}$ and $\mathbb{E}_{\mathcal{C}^{(i)}}(p_1^{(i)})$, as

$$\begin{aligned} & \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \{0,1\}^{n'}} \left| \mathbb{E}_{\mathcal{C}^{(i)}}(p_1^{(i)}(\vec{y}_w^{(i)})) - p_1^{(i)}(\vec{y}_w^{(i)}) \right| \\ & \leq \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \left| \sum_{\mathcal{C}^{(i)}} \Pr(\mathcal{C}^{(i)}) \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) - \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) \right| \end{aligned} \quad (32)$$

$$+ \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \left| \sum_{\mathcal{C}^{(i)}} \Pr(\mathcal{C}^{(i)}) \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) - \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) \right| \quad (33)$$

$$+ \frac{1}{2} \sum_{\vec{y}_w^{(i)} \notin \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \left| \sum_{\mathcal{C}^{(i)}} \Pr(\mathcal{C}^{(i)}) \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)}} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) - \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)}} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) \right| \quad (34)$$

$$\leq \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \left| \sum_{\mathcal{C}^{(i)}} \Pr(\mathcal{C}^{(i)}) \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) - \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) \right| \quad (35)$$

$$+ \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \mathbb{E}_{\mathcal{C}^{(i)}} \left(\sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) \right) + \frac{1}{2} \mathbb{E}_{\mathcal{C}^{(i)}} \left(\sum_{\vec{y}_w^{(i)} \notin \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}) \right) \quad (36)$$

$$+ \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) + \frac{1}{2} \sum_{\vec{y}_w^{(i)} \notin \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}). \quad (37)$$

The calculation above partitions the variational distance between the “actual” distribution $p_1^{(i)}$ and the “ensemble-averaged” distribution $\mathbb{E}_{\mathcal{C}}(p_1^{(i)})$ into three components. The term in (32) corresponds to the variational distance between $p_1^{(i)}$ and $\mathbb{E}_{\mathcal{C}}(p_1^{(i)})$ contributed by typical $\vec{y}_w^{(i)}$ and conditionally typical $\vec{x}^{(i)}$. The term in (33) corresponds to the variational distance contributed by typical $\vec{y}_w^{(i)}$ and conditionally atypical $\vec{x}^{(i)}$. The term in (34) corresponds to the variational distance contributed by atypical $\vec{y}_w^{(i)}$. Moreover, we bound (33) and (34) from above by the triangle inequality, and thus obtain the terms in (36) and (37). In the following, we will show that each term in (35), (36), (37) asymptotically goes to 0 (each term decreases faster than $\mathcal{O}(1/\sqrt{n})$) with high probability over inner code design.

Claim 5 (Term in (35)). *With probability (over inner code design) at least $1 - 2 \exp(-\frac{4}{3}n^{1/2+\delta/2})$,*

$$\frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}^{(i)})} \left| \sum_{\mathcal{C}^{(i)}} \Pr(\mathcal{C}^{(i)}) \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) - \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) \right| < n^{-1/2-\delta/4}.$$

Proof: We first calculate the probability (over inner code design) of one specific typical received vector $\vec{y}_w^{(i)}$ induced by conditionally typical $\vec{x}^{(i)}$.

$$\begin{aligned} & \sum_{\mathcal{C}^{(i)}} \Pr(\mathcal{C}^{(i)}) \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) \\ &= \sum_{\mathcal{C}^{(i)}} \Pr(\mathcal{C}^{(i)}) \sum_{f_{10}^w, f_{11}^w} \left(\sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{T}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})(f_{10}^w, f_{11}^w)} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) \right) \end{aligned} \quad (38)$$

$$= \sum_{\mathcal{C}^{(i)}} \Pr(\mathcal{C}^{(i)}) \sum_{f_{10}^w, f_{11}^w} \left| \mathcal{C}^{(i)} \cap \mathcal{T}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})(f_{10}^w, f_{11}^w) \right| p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) \quad (39)$$

$$= \sum_{f_{10}^w, f_{11}^w} \mathbb{E}_{\mathcal{C}^{(i)}} \left(\left| \mathcal{C}^{(i)} \cap \mathcal{T}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})(f_{10}^w, f_{11}^w) \right| \right) p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) \quad (40)$$

$$= \sum_{f_{10}^w, f_{11}^w} \Pr_{\mathcal{C}^{(i)}} \left(\vec{X}^{(i)} \in \mathcal{T}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})(f_{10}^w, f_{11}^w) \right) |\mathcal{C}^{(i)}| p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}). \quad (41)$$

To obtain equation (38), we decompose the conditionally typical set $\mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})$ into the conditional type classes $\mathcal{T}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})$ that comprise it. Equation (39) follows since $p(\vec{x}^{(i)})$ and $p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})$ are identical for all $\vec{x}^{(i)}$ in one conditional type class, and we interchange the order of the two summations to obtain equation (40). Equation (41) follows by noting that the expected number of inner codewords $\vec{x}^{(i)}$ in chunk i falling into a type class $\mathcal{T}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})$ equals the probability (over inner code design) of a single inner codeword in chunk i falling into the type class $\mathcal{T}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})$ times the size $|\mathcal{C}^{(i)}|$ of the inner codebook for chunk i . In the following, we bound from below the probability of a single inner-codeword falling into a specific type class.

$$\begin{aligned} & \Pr_{\mathcal{C}^{(i)}} \left(\vec{X}^{(i)} \in \mathcal{T}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})(f_{10}^w, f_{11}^w) \right) \\ &= \binom{n' (f_{01}^w + f_{11}^w)}{n' f_{11}^w} \rho^{n' f_{11}^w} (1 - \rho)^{n' f_{01}^w} \binom{n' (f_{00}^w + f_{10}^w)}{n' f_{10}^w} \rho^{n' f_{10}^w} (1 - \rho)^{n' f_{00}^w} \end{aligned} \quad (42)$$

$$\geq \frac{1}{2\pi k_1 k_2 \sqrt{p_w(1-p_w)} \log n} \cdot 2^{n' (f_{01}^w + f_{11}^w) H\left(\frac{f_{11}^w}{f_{01}^w + f_{11}^w}\right) + n' (f_{00}^w + f_{10}^w) H\left(\frac{f_{10}^w}{f_{00}^w + f_{10}^w}\right)} \rho^{n' (f_{10}^w + f_{11}^w)} (1 - \rho)^{n' (f_{00}^w + f_{01}^w)} \quad (43)$$

$$\begin{aligned} &= \frac{1}{2\pi k_1 k_2 \sqrt{p_w(1-p_w)} \log n} \cdot 2^{n' H(\vec{x}^{(i)}|\vec{y}_w^{(i)})} 2^{n' (f_{10}^w + f_{11}^w) \log \rho + n' (f_{00}^w + f_{01}^w) \log(1-\rho)} \\ &= \frac{1}{2\pi k_1 k_2 \sqrt{p_w(1-p_w)} \log n} \cdot 2^{n' H(\vec{x}^{(i)}|\vec{y}_w^{(i)})} \cdot 2^{-n' (H(\vec{x}^{(i)}) + D(\vec{x}^{(i)}\|\rho))} \end{aligned} \quad (44)$$

$$= \frac{1}{2\pi k_1 k_2 \sqrt{p_w(1-p_w)} \log n} \cdot 2^{-k_1 \sqrt{n} (\log n) [I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)}\|\rho)]} \quad (45)$$

$$= \frac{1}{2\pi k_1 k_2 \sqrt{p_w(1-p_w)} \log n} \cdot n^{-k_1 \sqrt{n} [I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)}\|\rho)]} \quad (46)$$

Equation (42) states the probability that $\vec{\mathbf{X}}^{(i)}$ satisfies the condition of one type class $\mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)}|\vec{y}_w^{(i)})$ given the received vector $\vec{y}_w^{(i)}$, based on standard counting arguments. In equation (43), we bound the binomial coefficients by the inequality

$$\binom{n}{k} \geq \left(\frac{1}{2\pi k}\right)^{1/2} 2^{nH(\frac{k}{n})},$$

which is derived from Stirling's approximation. The term $D(\vec{x}^{(i)} \parallel \rho)$ appearing in (44) is the empirical Kullback-Leibler divergence between $\vec{x}^{(i)}$ and the code design parameter ρ , defined in Section VII. In equation (45), we substitute the value of n' as $k_1\sqrt{n}(\log n)$, and merge the empirical entropy $H(\vec{x}^{(i)})$ and empirical conditional entropy $H(\vec{x}^{(i)}|\vec{y}_w^{(i)})$ into the empirical mutual information $I(\vec{x}^{(i)}; \vec{y}_w^{(i)})$. Recall that the empirical mutual information $I(\vec{x}^{(i)}; \vec{y}_w^{(i)})$ is a function of the triplet $(f_{*1}^w, f_{10}^w, f_{11}^w)$, and the empirical Kullback-Leibler divergence $D(\vec{x}^{(i)} \parallel \rho)$ is a function of the pair (f_{10}^w, f_{11}^w) . The range of f_{*1}^w, f_{10}^w , and f_{11}^w are the intervals $[\rho * p_w(1 \pm \Delta_{*1}^w)]$, $[\rho p_w(1 \pm \Delta_{10}^w)]$, and $[\rho(1 - p_w)(1 \pm \Delta_{11}^w)]$ respectively since we only consider typical $\vec{y}_w^{(i)}$ and the conditionally typical $\vec{x}^{(i)}$ here.

To figure out the value of $(f_{*1}^w, f_{10}^w, f_{11}^w)$ that maximizes $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$, we take partial derivatives of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)})$ and $D(\vec{x}^{(i)} \parallel \rho)$ with respect to f_{*1}^w, f_{10}^w and f_{11}^w in Appendix D. It turns out that for different value of p_w , the maximal value of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ is attained at different points. Though we are unable to give a specific value $(f_{*1}^w, f_{10}^w, f_{11}^w)$ that maximizes $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$, in Appendix D we can still make sure that the maximum is attained at one of the four points, i.e., $f_{*1}^w = \rho * p_w, f_{10}^w = \rho p_w(1 \pm \Delta_{10}^w)$ and $f_{11}^w = \rho(1 - p_w)(1 \pm \Delta_{11}^w)$. In Appendix E, we show that when the blocklength n is sufficiently large,

$$\lim_{n \rightarrow \infty} -k_1\sqrt{n} \left(I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho) \right) \geq k_1 \max_{i=1}^4 \{g_i(p_w, \epsilon_d, \Delta_{10}^w, \Delta_{11}^w)\}, \quad (47)$$

where the auxiliary functions $g_i(\cdot, \cdot, \cdot, \cdot)$ is defined in (4)-(7), Section IV. Recall that as specified in Section V-A, the size of the codebook $\mathcal{C}^{(i)}$ is $n^{k_1 r_u}$. Hence substituting (46) into (41) gives us that the expected number of inner codewords $\vec{x}^{(i)}$ falling into the type-class $\mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)}|\vec{y}_w^{(i)})$ equals

$$\begin{aligned} \mathbb{E}_{\mathcal{C}^{(i)}} \left(\left| \mathcal{C}^{(i)} \cap \mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)}|\vec{y}_w^{(i)})(f_{10}^w, f_{11}^w) \right| \right) &= \Pr_{W^{(i)}} \left(X^{(i)} \in \mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)}|\vec{y}_w^{(i)})(f_{10}^w, f_{11}^w) \right) \cdot |\mathcal{C}^{(i)}| \\ &\geq n^{k_1 r_u + k_1 \max_{i=1}^4 g_i(p_w, \epsilon_d, \Delta_{10}^w, \Delta_{11}^w)}. \end{aligned} \quad (48)$$

As claimed in Section IV, we set the code chunk length design parameter k_1 such that $k_1 r_u + k_1 \max_{i=1}^4 g_i(p_w, \epsilon_d, \Delta_{10}^w, \Delta_{11}^w) \geq 3/2 + \delta$. Therefore, we have

$$\mathbb{E}_{\mathcal{C}^{(i)}} \left(\left| \mathcal{C}^{(i)} \cap \mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)}|\vec{y}_w^{(i)})(f_{10}^w, f_{11}^w) \right| \right) \geq n^{3/2+\delta}. \quad (49)$$

For notational convenience, we denote the number of codewords falling into one type class, i.e., $\left| \mathcal{C}^{(i)} \cap \mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)}|\vec{y}_w^{(i)})(f_{10}^w, f_{11}^w) \right|$, by $\mathcal{R}(\mathcal{C}^{(i)}, \vec{y}_w^{(i)}, f_{10}^w, f_{11}^w)$, and hence $\mathbb{E}_{\mathcal{C}^{(i)}}(\mathcal{R}(\mathcal{C}^{(i)}, \vec{y}_w^{(i)}, f_{10}^w, f_{11}^w)) \geq n^{3/2+\delta}$. By the Chernoff bound¹¹ [35], we obtain that the actual number of codewords falling into one type class is tightly concentrated around its expectation, i.e.,

$$\begin{aligned} &\Pr_{\mathcal{C}^{(i)}, W^{(i)}} \left\{ \left| \mathcal{R}(\mathcal{C}^{(i)}, \vec{y}_w^{(i)}, f_{10}^w, f_{11}^w) - \mathbb{E}_{\mathcal{C}^{(i)}}(\mathcal{R}(\mathcal{C}^{(i)}, \vec{y}_w^{(i)}, f_{10}^w, f_{11}^w)) \right| > 2n^{-1/2-\delta/4} \mathbb{E}_{\mathcal{C}^{(i)}}(\mathcal{R}(\mathcal{C}^{(i)}, \vec{y}_w^{(i)}, f_{10}^w, f_{11}^w)) \right\} \\ &< 2 \exp \left(-\frac{4}{3} n^{-1-\delta/2} n^{\frac{3}{2}+\delta} \right) = 2 \exp \left(-\frac{4}{3} n^{1/2+\delta/2} \right). \end{aligned} \quad (50)$$

If we only focus on one side of the concentration inequality (50), it then follows that with probability (over inner code design) at most $\exp(-\frac{4}{3} n^{1/2+\delta/2})$, the number of codewords falling into one type class is less than

$$(1 - 2n^{-1/2-\delta/4}) \mathbb{E}_{\mathcal{C}^{(i)}}(\mathcal{R}(\mathcal{C}^{(i)}, \vec{y}_w^{(i)}, f_{10}^w, f_{11}^w)) = n^{3/2+\delta} - n^{1+3\delta/4}.$$

Since $n^{3/2}$ is smaller than $n^{3/2+\delta} - n^{1+3\delta/4}$, we then have the following claim.

Claim 6. *With probability (over inner code design) at least $1 - \exp(-\frac{4}{3} n^{1/2+\delta/2})$, for a typical $\vec{y}_w^{(i)}$, there are at least $n^{3/2}$ codewords $\vec{x}^{(i)}$ falling into the conditionally typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)}|\vec{y}_w^{(i)})$.*

¹¹We state the version of the Chernoff bound we used here (and throughout this paper) in Appendix A, since there are many different versions of the Chernoff bound in the literature.

Returning now to estimating the term in (35), we thus conclude that with probability (over inner code design) at least $1 - 2 \exp(-\frac{4}{3}n^{1/2+\delta/2})$,

$$\begin{aligned} & \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} \left| \sum_{\mathcal{C}^{(i)}} \Pr(\mathcal{C}^{(i)}) \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) - \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \right| \\ & \leq \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} \sum_{f_{10}^w, f_{11}^w} \left| \mathbb{E}_{\mathcal{C}^{(i)}} \left(\mathcal{R}(\mathcal{C}^{(i)}, \vec{y}_w^{(i)}, f_{10}^w, f_{11}^w) \right) - \mathcal{R}(\mathcal{C}^{(i)}, \vec{y}_w^{(i)}, f_{10}^w, f_{11}^w) \right| p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \end{aligned} \quad (51)$$

$$< \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} \sum_{f_{10}^w, f_{11}^w} 2n^{-1/2-\delta/4} \mathbb{E}_{\mathcal{C}^{(i)}} \left(\mathcal{R}(\mathcal{C}^{(i)}, \vec{y}_w^{(i)}, f_{10}^w, f_{11}^w) \right) p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \quad (52)$$

$$< n^{-1/2-\delta/4}. \quad (53)$$

Analogously to the decomposition in equations (38)-(41), to obtain equation (51), we decompose the conditionally typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_w^{(i)})$ into the summation over all the conditional type class $\mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_w^{(i)})(f_{10}^w, f_{11}^w)$. Equation (52) follows from the fact, stated in (50), that the number of codewords falling into one conditional type class is tightly concentrated around its expectation. Equation (53) holds since

$$\begin{aligned} & \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} \sum_{f_{10}^w, f_{11}^w} \mathbb{E}_{\mathcal{C}^{(i)}} \left(\mathcal{R}(\mathcal{C}^{(i)}, \vec{y}_w^{(i)}, f_{10}^w, f_{11}^w) \right) p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \\ & = \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} \mathbb{E}_{\mathcal{C}^{(i)}} \left(\mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_w^{(i)}) \right) p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \\ & \leq \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)}} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \\ & \leq \sum_{\vec{y}_w^{(i)}} \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)}} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \\ & = 1. \end{aligned}$$

This completes the proof of Claim 5. \square

In the following, we show that from Willie's perspective the probability (over inner code design) of receiving an atypical $\vec{y}_w^{(i)}$ goes to 0 asymptotically. We choose Δ_{*1}^w as $n^{-1/4+\delta/2}$ (recall that Δ_{*1}^w is the parameter, defined in Section VII, specifying the "width" of the narrow typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})$).

Claim 7 (Term in (36)). *The probability (over inner code design) of receiving an atypical $\vec{y}_w^{(i)}$ is bounded from above as*

$$\sum_{\vec{y}_w^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} \mathbb{E}_{\mathcal{C}^{(i)}} \left(p_1^{(i)}(\vec{y}_w^{(i)}) \right) < 2n^{-\frac{1}{3}k_1(\rho * p_w)n^\delta \log e}.$$

Proof: Note that the ensemble average distribution $\mathbb{E}_{\mathcal{C}^{(i)}}(p_1^{(i)}(\vec{y}_w^{(i)}))$ is a Bernoulli($p_w * \rho$) distribution, since it corresponds to an inner codeword $\vec{x}^{(i)}$ being chosen according to a Bernoulli(ρ) distribution, and then $\vec{x}^{(i)}$ passing through a BSC(p_w). The probability that a $\vec{y}_w^{(i)}$ generated in this manner is atypical (the type-class f_{*1}^w falls outside the range $[(1 \pm \Delta_{*1}^w)\rho * p_w]$) is at most $n^{-\frac{1}{3}k_1(\rho * p_w)n^\delta \log e}$ by the Chernoff bound, since the value of Δ_{*1}^w is chosen as $n^{-1/4+\delta/2}$. More specifically, we have

$$\begin{aligned} \sum_{\vec{y}_w^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} \mathbb{E}_{\mathcal{C}^{(i)}} \left(p_1^{(i)}(\vec{y}_w^{(i)}) \right) & = \Pr_{\mathcal{C}^{(i)}, W^{(i)}, \vec{\mathbf{Z}}_w} \left(\vec{y}_w^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)}) \right) \\ & = \Pr_{\mathcal{C}^{(i)}, W^{(i)}, \vec{\mathbf{Z}}_w} \left(f_{*1}^w \notin [(1 \pm \Delta_{*1}^w)\rho * p_w] \right) \\ & < 2 \exp \left(-\frac{1}{3} (\Delta_{*1}^w)^2 (\rho * p_w) k_1 \sqrt{n} \log n \right) \\ & = 2n^{-\frac{1}{3}k_1(\rho * p_w)n^\delta \log e} \end{aligned} \quad (54)$$

The inequality (54) follows from the Chernoff bound, as mentioned above. \square

In the following claim, we show that the probability (over inner code design) of receiving a typical $\vec{y}_w^{(i)}$ if a conditionally atypical codeword $\vec{x}^{(i)}$ is transmitted is polynomially small.

Claim 8 (Term in (36)). *The probability (over inner code design) of receiving a typical $\vec{y}_w^{(i)}$ if a conditionally atypical codeword $\vec{x}^{(i)}$ is transmitted is bounded from above as*

$$\mathbb{E}_{\mathcal{C}^{(i)}} \left(\sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) \right) < 4 \cdot n^{-1/2-\delta}.$$

Proof:

$$\begin{aligned} & \mathbb{E}_{\mathcal{C}^{(i)}} \left(\sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}|\vec{x}^{(i)})p(\vec{x}^{(i)}) \right) \\ &= \mathbb{E}_{\mathcal{C}^{(i)}} \left(\sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)})p(\vec{x}^{(i)}|\vec{y}_w^{(i)}) \right) \\ &= \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}) \mathbb{E}_{\mathcal{C}^{(i)}} \left(\sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p(\vec{x}^{(i)}|\vec{y}_w^{(i)}) \right) \end{aligned} \quad (55)$$

$$\begin{aligned} & \leq \mathbb{E}_{\mathcal{C}^{(i)}} \left(\sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{X}^{(i)}|\vec{y}_w^{(i)})} p(\vec{x}^{(i)}|\vec{y}_w^{(i)}) \right) \\ &= \Pr_{\mathcal{C}^{(i)}, W^{(i)}, \vec{Z}_w} \left(f_{10}^w \notin [(1-\Delta_{10}^w)\rho p_w, (1+\Delta_{10}^w)\rho p_w] \cup f_{11}^w \notin [(1-\Delta_{11}^w)\rho(1-p_w), (1+\Delta_{11}^w)\rho(1-p_w)] \right), \end{aligned} \quad (56)$$

where equation (55) is obtained by interchanging the order of summation, and equation (56) follows since the probability of receiving a typical $\vec{y}_w^{(i)}$ is smaller than one. Then we use standard counting arguments to obtain

$$\begin{aligned} & \Pr_{\mathcal{C}^{(i)}, W^{(i)}, \vec{Z}_w} \left(f_{10}^w \notin [(1-\Delta_{10}^w)\rho p_w, (1+\Delta_{10}^w)\rho p_w] \cup f_{11}^w \notin [(1-\Delta_{11}^w)\rho(1-p_w), (1+\Delta_{11}^w)\rho(1-p_w)] \right) \\ & \leq \sum_{i_1=k_1 k_2 p_w (\log n)(1+\Delta_{10}^w)}^{k_1 \sqrt{n} \log n} \binom{k_1 \sqrt{n} \log n}{i_1} \left(\frac{k_2 p_w}{\sqrt{n}} \right)^{i_1} \left(1 - \frac{k_2 p_w}{\sqrt{n}} \right)^{k_1 \sqrt{n} \log n - i_1} \end{aligned} \quad (57)$$

$$+ \sum_{i_2=0}^{k_1 k_2 p_w (\log n)(1-\Delta_{10}^w)} \binom{k_1 \sqrt{n} \log n}{i_2} \left(\frac{k_2 p_w}{\sqrt{n}} \right)^{i_2} \left(1 - \frac{k_2 p_w}{\sqrt{n}} \right)^{k_1 \sqrt{n} \log n - i_2} \quad (58)$$

$$+ \sum_{i_3=k_1 k_2 (1-p_w)(\log n)(1+\Delta_{11}^w)}^{k_1 \sqrt{n} \log n} \binom{k_1 \sqrt{n} \log n}{i_3} \left(\frac{k_2 (1-p_w)}{\sqrt{n}} \right)^{i_3} \left(1 - \frac{k_2 (1-p_w)}{\sqrt{n}} \right)^{k_1 \sqrt{n} \log n - i_3} \quad (59)$$

$$+ \sum_{i_4=0}^{k_1 k_2 (1-p_w)(\log n)(1-\Delta_{11}^w)} \binom{k_1 \sqrt{n} \log n}{i_4} \left(\frac{k_2 (1-p_w)}{\sqrt{n}} \right)^{i_4} \left(1 - \frac{k_2 (1-p_w)}{\sqrt{n}} \right)^{k_1 \sqrt{n} \log n - i_4} \quad (60)$$

Here the four terms in (57)-(60) correspond to the four possible atypical ranges for the pair (f_{10}^w, f_{11}^w) . In Appendix B, we show that term (57) and term (59) respectively satisfy

$$\sum_{i_1=k_1 k_2 p_w (\log n)(1+\Delta_{10}^w)}^{k_1 \sqrt{n} \log n} \binom{k_1 \sqrt{n} \log n}{i_1} \left(\frac{k_2 p_w}{\sqrt{n}} \right)^{i_1} \left(1 - \frac{k_2 p_w}{\sqrt{n}} \right)^{k_1 \sqrt{n} \log n - i_1} \leq n^{-k_1 k_2 p_w f(\Delta_{10}^w)} \quad (61)$$

$$\sum_{i_3=k_1 k_2 (1-p_w)(\log n)(1+\Delta_{11}^w)}^{k_1 \sqrt{n} \log n} \binom{k_1 \sqrt{n} \log n}{i_3} \left(\frac{k_2 (1-p_w)}{\sqrt{n}} \right)^{i_3} \left(1 - \frac{k_2 (1-p_w)}{\sqrt{n}} \right)^{k_1 \sqrt{n} \log n - i_3} \leq n^{-k_1 k_2 (1-p_w) f(\Delta_{11}^w)}, \quad (62)$$

where the auxiliary function $f(\cdot)$ is as defined in Section IV, and term (58) and term (60) are strictly smaller than term (57) and term (59) respectively. Recall that in Section IV, the code chunk length design parameter k_1 satisfies the following two conditions¹²:

$$\begin{cases} k_1 k_2 p_w \cdot f(\Delta_{10}^w) \geq 1/2 + \delta, \\ k_1 k_2 (1 - p_w) \cdot f(\Delta_{11}^w) \geq 1/2 + \delta. \end{cases} \quad (63)$$

Therefore, we have

$$\begin{aligned} \mathbb{E}_{\mathcal{C}^{(i)}} \left(\sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \right) &\leq (2n^{-1/2-\delta} + 2n^{-1/2-\delta}) \\ &= 4 \cdot n^{-1/2-\delta}. \end{aligned}$$

□

In Claim 9, we show that for a randomly chosen inner code $\mathcal{C}^{(i)}$, the probability of receiving an atypical $\vec{y}_w^{(i)}$ plus the probability of receiving a typical $\vec{y}_w^{(i)}$ induced by a conditionally atypical codeword $\vec{x}^{(i)}$ is polynomially small.

Claim 9 (Term in (37)). *For a randomly chosen inner code $\mathcal{C}^{(i)}$, with probability (over inner code design) at least $1 - 2 \exp(-\frac{4}{3}n^{1/2+\delta/2})$,*

$$\frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) + \frac{1}{2} \sum_{\vec{y}_w^{(i)} \notin \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}) < 4 \cdot n^{-1/2-\delta}.$$

Proof: By combining Claim 5 and Claim 8, with probability (over inner code design) at least $1 - 2 \exp(-\frac{4}{3}n^{1/2+\delta/2})$, we have

$$\begin{aligned} &n^{-1/2-\delta/4} + 2 \cdot n^{-1/2-\delta} \\ &> \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \left| \sum_{\mathcal{C}^{(i)}} \Pr(\mathcal{C}^{(i)}) \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) - \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \right| \\ &+ \frac{1}{2} \mathbb{E}_{\mathcal{C}^{(i)}} \left(\sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \left(\sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \right) \right) \\ &\geq \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \left| \sum_{\mathcal{C}^{(i)}} \Pr(\mathcal{C}^{(i)}) \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \right| \\ &+ \frac{1}{2} \mathbb{E}_{\mathcal{C}^{(i)}} \left(\sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \left(\sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \right) \right) \\ &- \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \left(\sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \right), \end{aligned} \quad (65)$$

$$\begin{aligned} &- \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \left(\sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \right), \end{aligned} \quad (66)$$

where equation (66) follows from the triangle inequality. Note that the summation of the first two terms of (66) equals the probability of typical $\vec{y}_w^{(i)}$ under the ensemble average distribution, which, by Claim 7, equals

$$\frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \mathbb{E}_{\mathcal{C}^{(i)}} \left(p_1^{(i)}(\vec{y}_w^{(i)}) \right) = 1 - \frac{1}{2} \sum_{\vec{y}_w^{(i)} \notin \mathcal{A}_{n'}^1(\vec{Y}_w^{(i)})} \mathbb{E}_{\mathcal{C}^{(i)}} \left(p_1^{(i)}(\vec{y}_w^{(i)}) \right) = 1 - n^{-\frac{1}{3}k_1(\rho^* p_w) n^\delta \log e}. \quad (67)$$

¹²In order to show the probability (over inner code design) of receiving a typical $\vec{y}_w^{(i)}$ if a conditionally atypical codeword $\vec{x}^{(i)}$ is transmitted is bounded from above by $\mathcal{O}(n^{-1/2-\delta})$, we require each of the four terms in (57)-(60) to scale as $\mathcal{O}(n^{-1/2-\delta})$. And this requirement, in turn, forces us to scale Δ_{10}^w and Δ_{11}^w as constants (in the interval $[0, 1]$) that satisfy the constraints in (63) and (64).

For the third term of (66), we have

$$\begin{aligned} & \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} \left(\sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \right) \\ &= 1 - \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} \left(\sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \right) - \frac{1}{2} \sum_{\vec{y}_w^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}). \end{aligned} \quad (68)$$

Hence, combining equations (66), (67) and (68), with probability (over inner code design) at least $1 - 2 \exp(-\frac{4}{3}n^{1/2+\delta/2})$ we have

$$\begin{aligned} & \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} \left(\sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \right) + \frac{1}{2} \sum_{\vec{y}_w^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}) \\ &< n^{-1/2-\delta/4} + 2 \cdot n^{-1/2-\delta} + n^{-\frac{1}{3}k_1(\rho * p_w)n^\delta \log e} \\ &< 4 \cdot n^{-1/2-\delta}. \end{aligned}$$

This completes the proof of Claim 9. \square

Equipped with Claims 5-9, we are able to show that for any chunk $i \in \{1, \dots, L\}$, the ensemble-averaged distribution $\mathbb{E}_{\mathcal{C}^{(i)}}(p_1^{(i)}(\vec{y}_w^{(i)}))$ and the actual distribution $p_1^{(i)}(\vec{y}_w^{(i)})$ are pretty close with high probability over inner code design.

Claim 10. *For any $i \in \{1, \dots, L\}$, with probability (over inner code design) at least $1 - 2 \exp(-\frac{4}{3}n^{1/2+\delta/2})$, the ensemble-averaged distribution $\mathbb{E}_{\mathcal{C}^{(i)}}(p_1^{(i)}(\vec{y}_w^{(i)}))$ is close to the actual distribution $p_1^{(i)}(\vec{y}_w^{(i)})$, i.e.,*

$$\frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \{0,1\}^{n'}} \left| \mathbb{E}_{\mathcal{C}^{(i)}}(p_1^{(i)}(\vec{y}_w^{(i)})) - p_1^{(i)}(\vec{y}_w^{(i)}) \right| < 8 \cdot n^{-1/2-\delta/4}.$$

Proof: For any $i \in \{1, \dots, L\}$, with probability (over inner code design) at least $1 - 2 \exp(-\frac{4}{3}n^{1/2+\delta/2})$,

$$\begin{aligned} & \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \{0,1\}^{n'}} \left| \mathbb{E}_{\mathcal{C}^{(i)}}(p_1^{(i)}(\vec{y}_w^{(i)})) - p_1^{(i)}(\vec{y}_w^{(i)}) \right| \\ &\leq \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} \left| \sum_{\mathcal{C}^{(i)}} \Pr(\mathcal{C}^{(i)}) \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) - \sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \cap \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \right| \end{aligned} \quad (69)$$

$$+ \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} \mathbb{E}_{\mathcal{C}^{(i)}} \left(\sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \right) + \frac{1}{2} \mathbb{E}_{\mathcal{C}^{(i)}} \left(\sum_{\vec{y}_w^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}) \right) \quad (70)$$

$$+ \frac{1}{2} \sum_{\vec{y}_w^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} \left(\sum_{\vec{x}^{(i)} \in \mathcal{C}^{(i)} \setminus \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)} | \vec{x}^{(i)}) p(\vec{x}^{(i)}) \right) + \frac{1}{2} \sum_{\vec{y}_w^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_w^{(i)})} p_1^{(i)}(\vec{y}_w^{(i)}) \quad (71)$$

$$\begin{aligned} &< n^{-1/2-\delta/4} + n^{-\frac{1}{3}k_1(\rho * p_w)n^\delta \log e} + 2 \cdot n^{-1/2-\delta} + 4 \cdot n^{-1/2-\delta} \\ &< 8 \cdot n^{-1/2-\delta/4}, \end{aligned} \quad (72)$$

where inequalities (69)-(71) are adapted from (32)-(34), and inequality (72) follows from Claims 5-9. \square

In the following, we take one more step to show that with high probability over concatenated code design, the n -letter ensemble-averaged distribution and the “chunk-wise independent” distribution are close.

Lemma 3 (Restated). *With probability at least $1 - \sqrt{n} \exp(-\frac{4}{3}n^{1/2+\delta/2})$ over channel noise to Willie and concatenated code design, the ensemble-averaged distribution is close to the “chunk-wise independent” product distribution, i.e.,*

$$\frac{1}{2} \sum_{\vec{y}_w^{(1)} \dots \vec{y}_w^{(L)}} \left| \mathbb{E}_{\mathcal{C}}(p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(L)})) - p_1^{(1)}(\vec{y}_w^{(1)}) \dots p_1^{(L)}(\vec{y}_w^{(L)}) \right| < n^{-\delta/4}.$$

Proof: Based on Claim 10 and the union bound, it is then the case that with probability at least $1 - L \cdot 2 \exp(-\frac{4}{3}n^{1/2+\delta/2})$ over concatenated code design, the variational distance between the ensemble-averaged distribution and “chunk-wise independent” product distribution is bounded from above as

$$\begin{aligned}
& \frac{1}{2} \sum_{\vec{y}_w^{(1)} \dots \vec{y}_w^{(L)}} \left| \mathbb{E}_{\mathcal{C}}(p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(L)})) - p_1^{(1)}(\vec{y}_w^{(1)}) \dots p_1^{(L)}(\vec{y}_w^{(L)}) \right| \\
& \leq \frac{1}{2} \sum_{i=1}^L \sum_{\vec{y}_w^{(i)} \in \{0,1\}^{n'}} \left| \mathbb{E}_{\mathcal{C}^{(i)}}(p_1^{(i)}(\vec{y}_w^{(i)})) - p_1^{(i)}(\vec{y}_w^{(i)}) \right| \\
& = 8L \cdot n^{-1/2-\delta/4} \\
& = \frac{8n^{1/2}}{k_1 \log n} \cdot n^{-1/2-\delta/4} \\
& \leq n^{-\delta/4},
\end{aligned}$$

for sufficiently large n . Note that $1 - L \cdot 2 \exp(-\frac{4}{3}n^{1/2+\delta/2}) \leq 1 - \sqrt{n} \exp(-\frac{4}{3}n^{1/2+\delta/2})$, since $L = \sqrt{n}/(k_1 \log n)$. This completes the proof of Lemma 3. \square

B. Proof of Lemma 4:

We note that $p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(\lambda L)}) = p_1^{(1)}(\vec{y}_w^{(1)}) \dots p_1^{(\lambda L)}(\vec{y}_w^{(\lambda L)})$ since the inner codes in the first λL systematic chunks, and also the messages $\mathbf{W}^{(i)} = \mathbf{M}^{(i)}$ that are inputs to those chunks, are all independent. However, the analysis for the remaining $L(1-\lambda)$ parity chunks is more involved since the Reed-Solomon outer code in general introduces correlations between $\mathbf{W}^{(i)}$ in the λL systematic chunks (which are λL -wise independent) and any $\mathbf{W}^{(i')}$ in a parity chunk – in particular, any such $\mathbf{W}^{(i')}$ is a linear combination of $\mathbf{W}^{(i)}$ in the λL systematic chunks.

Let $l_1 = \lambda L$ be the number of systematic chunks and $l_2 = L(1-\lambda)$ be the number of parity chunks. Next, we intend to show that from Willie’s perspective, the inner-messages $\mathbf{W}^{(\lambda L+1)}, \dots, \mathbf{W}^{(L)}$ of all parity chunks are almost uniformly distributed (essentially statistically independent of observed transmissions in the systematic chunks). As noted earlier, the inner messages $\mathbf{W}^{(i)}$ for systematic chunks can be directly seen to be independently and uniformly distributed since for such chunks $\mathbf{W}^{(i)}$, and the messages $\mathbf{M}^{(i)}$ for systematic chunks are independently and uniformly distributed.

Claim 11. *From Willie’s perspective, the l_2 -tuple $(\mathbf{W}^{(\lambda L+1)}, \dots, \mathbf{W}^{(L)})$ is almost uniformly distributed, i.e., $\frac{1}{|W^{(L)}|^{l_2}}(1-n^{-1}) \leq p(\mathbf{W}^{(\lambda L+1)} = w^{(\lambda L+1)}, \dots, \mathbf{W}^{(L)} = w^{(L)} | \vec{y}_w^{(1)}, \dots, \vec{y}_w^{(\lambda L)}) \leq \frac{1}{|W^{(L)}|^{l_2}}(1+n^{-1})$, for all l_2 -tuples $(w^{(\lambda L+1)}, \dots, w^{(L)})$, with probability at least $1 - \mathcal{O}(n^{-\delta}/\log n)$ over channel noise to Willie and concatenated code design.*

Proof: We assume an oracle reveals to Willie which conditional type class Alice’s codeword is in¹³. Claim 7 states that for any $i \in \{1, \dots, L\}$, the probability (over inner code design) of receiving a typical $\vec{y}_w^{(i)}$ is at least $1 - \exp(-\mathcal{O}(n^\delta \log n))$. Claim 8 shows that with probability at least $1 - 4 \cdot n^{-1/2-\delta}$, a typical $\vec{y}_w^{(i)}$ is induced by a typical codeword $\vec{x}^{(i)}$. Conditioned on the fact that Willie receives a typical $\vec{y}_w^{(i)}$ which is induced by a typical $\vec{x}^{(i)}$, we then note that with probability at least $1 - \exp(-\mathcal{O}(\sqrt{n}))$, the number of inner codewords $\vec{x}^{(i)}$ in the revealed chunk-wise conditional type class $\mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_w^{(i)})(f_{10}^w, f_{11}^w)$ within the chunk-wise conditionally typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_w^{(i)})$, is at least $n^{3/2}$ (as proved in Claim 6), and each such $\vec{x}^{(i)}$ has the same probability $p(\vec{x}^{(i)} | \vec{y}_w^{(i)}, \text{oracle})$. Combining the analysis above together, we conclude that the probability (over inner code design and channel noise to Willie) that the revealed conditional type class contains at least $n^{3/2}$ codewords, is bounded from below as

$$(1 - \exp(-\mathcal{O}(n^\delta \log n))) \cdot (1 - 4 \cdot n^{-1/2-\delta}) \cdot (1 - \exp(-\mathcal{O}(\sqrt{n}))) \geq 1 - 6 \cdot n^{-1/2-\delta},$$

for all sufficiently large n . Taking a union bound over all of the $L = \sqrt{n}/(k_1 \log n)$ chunks, we obtain that with probability (over inner code design and channel noise to Willie) at least $1 - \mathcal{O}(n^{-\delta}/\log n)$, all of the L revealed conditional type classes contains at least $n^{3/2}$ codewords.

Based on the analysis above, it then follows that the number of possible values that the systematic inner-message vector $(\mathbf{W}^{(1)}, \dots, \mathbf{W}^{\lambda L})$ can take, conditioned on the oracle revealing to Willie which conditional type class each $\mathbf{W}^{(i)}$ lies in w.r.t. each $\vec{y}_w^{(i)}$, is at least $(n^{3/2})^{l_1}$ since the encoding of each systematic chunk is independent of every other chunk. In the following, we call one possible value of the systematic inner-message vector $(\mathbf{W}^{(1)}, \dots, \mathbf{W}^{\lambda L})$ as *one combination* for convenience. Let the number of combinations be n^ν (the random variable ν exceeds $3l_1/2$ with high probability over concatenated code design),

¹³The reason for introducing such an oracle is that then each plausible codeword is of equal probability from Willie’s perspective, if the true conditional type class is revealed to Willie. It is not really necessary to use this trick – it merely simplifies calculations in Claim 11.

and for any $1 \leq i \leq n^\nu$, we denote the i -th combination by S_i . Note that the inner-message parity vector $(\mathbf{W}^{(1)}, \dots, \mathbf{W}^{(L)})$ is a deterministic function of the systematic inner message $(\mathbf{W}^{(1)}, \dots, \mathbf{W}^{(L)})$, under the action of the Reed-Solomon outer code. The properties of Reed-Solomon codes ensure that (as proved in Appendix C) each inner-message parity vector corresponds to an approximately equal number of systematic inner messages (combinations). More specifically, given a specific parity inner-message vector $(w^{(\lambda L+1)}, \dots, w^{(L)})$, the number of combinations that could have caused it is $n^{\nu-l_2 k_1 r}$ on average, since the number of parity chunks is l_2 and each chunk contains $n^{k_1 r}$ inner-messages. We define U as the number of combinations leading to a specific parity inner-message vector, say $(w_1^{(\lambda L+1)}, w_1^{(\lambda L+2)}, \dots, w_1^{(L)})$. As mentioned earlier, there are n^ν combinations of systematic chunks. For each combination S_i ($1 \leq i \leq n^\nu$), we define the indicator function

$$\mathbb{1}(S_i) = \begin{cases} 1, & \text{if } S_i \text{ leads to } (w_1^{(\lambda L+1)}, w_1^{(\lambda L+2)}, \dots, w_1^{(L)}) \\ 0, & \text{otherwise.} \end{cases}$$

In the following, we calculate $\mathbb{E}(U)$ and $\mathbb{E}(U^2)$,

$$\begin{aligned} \mathbb{E}(U) &= \mathbb{E} \left[\sum_{i=1}^{n^\nu} \mathbb{1}(S_i) \right] = n^{\nu-l_2 k_1 r} \\ \mathbb{E}(U^2) &= \mathbb{E} \left[\left(\sum_{i=1}^{n^\nu} \mathbb{1}(S_i) \right)^2 \right] \\ &= \mathbb{E} \left[\sum_{i=1}^{n^\nu} \mathbb{1}(S_i)^2 + \sum_{i \neq j} \mathbb{1}(S_i) \cdot \mathbb{1}(S_j) \right] \\ &= \mathbb{E} \left[\sum_{i=1}^{n^\nu} \mathbb{1}(S_i)^2 \right] + \sum_{i \neq j} \Pr[(\mathbb{1}(S_i) = 1) \cap (\mathbb{1}(S_j) = 1)] \\ &= \mathbb{E}(U) + \sum_{i \neq j} \Pr[(\mathbb{1}(S_i) = 1) \cap (\mathbb{1}(S_j) = 1)] \end{aligned}$$

For the latter term $\Pr[(\mathbb{1}(S_i) = 1) \cap (\mathbb{1}(S_j) = 1)]$, the two combinations S_i and S_j are sampled without replacement. If we consider sampling with replacement, then the probability that two combinations lead to a specific tuple equals

$$\begin{aligned} \Pr[(\mathbb{1}(S_i) = 1) \cap (\mathbb{1}(S_j) = 1) \text{ with replacement}] &= \Pr[\mathbb{1}(S_i) = 1] \cdot \Pr[\mathbb{1}(S_j) = 1] \\ &= \frac{1}{n^{l_2 k_1 r}} \cdot \frac{1}{n^{l_2 k_1 r}}. \end{aligned}$$

Intuitively, the probability of $\mathbb{1}(S_i) = 1$ and $\mathbb{1}(S_j) = 1$ happening simultaneously will decrease if the two combinations are sampled without replacement. We show that this is true in the following,

$$\begin{aligned} \Pr[(\mathbb{1}(S_i) = 1) \cap (\mathbb{1}(S_j) = 1)] &= \sum_u \Pr(U = u) \Pr[(\mathbb{1}(S_i) = 1) \cap (\mathbb{1}(S_j) = 1) | U = u] \\ &= \sum_u \Pr(U = u) \frac{u}{n^\nu} \frac{u-1}{n^\nu-1} \\ &\leq \sum_u \Pr(U = u) \frac{u}{n^\nu} \frac{u}{n^\nu} \\ &= \sum_u \Pr(U = u) \Pr[(\mathbb{1}(S_i) = 1) \cap (\mathbb{1}(S_j) = 1) \text{ with replacement} | U = u] \\ &= \Pr[(\mathbb{1}(S_i) = 1) \cap (\mathbb{1}(S_j) = 1) \text{ with replacement}] \\ &= \frac{1}{n^{l_2 k_1 r}} \cdot \frac{1}{n^{l_2 k_1 r}}. \end{aligned}$$

By combining the two steps above, we obtain

$$\begin{aligned} \mathbb{E}(U^2) &= \mathbb{E}(U) + \sum_{i \neq j} \Pr[(\mathbb{1}(S_i) = 1) \cap (\mathbb{1}(S_j) = 1)] \\ &\leq n^{\nu-l_2 k_1 r} + (n^{2\nu} - n^\nu) (n^{-2l_2 k_1 r}). \end{aligned}$$

We now use the *second-moment method* to bound from above the probability that U is significantly smaller than its expectation. To this end it is useful to define the random variable V as $U - (1 - \varepsilon)\mathbb{E}(U)$, where ε is a parameter with value to be specified

later (at the end of Claim 11). Note that the expected value of V is given as

$$\mathbb{E}(V) = \varepsilon \mathbb{E}(U) = \varepsilon n^{\nu-l_2 k_1 r}. \quad (73)$$

Moreover, note that

$$\begin{aligned} \mathbb{E}(V^2) &= \mathbb{E} \left(U^2 + (1-\varepsilon)^2 (\mathbb{E}(U))^2 - 2(1-\varepsilon)U \cdot \mathbb{E}(U) \right) \\ &= \mathbb{E}(U^2) + (\varepsilon^2 - 1) (\mathbb{E}(U))^2 \\ &\leq n^{\nu-l_2 k_1 r} - n^{\nu-2l_2 k_1 r} + \varepsilon^2 n^{2\nu-2l_2 k_1 r} \end{aligned} \quad (74)$$

We now introduce an auxiliary random variable V' defined as follows:

$$V' = \begin{cases} U - (1-\varepsilon)\mathbb{E}(U), & \text{if } U - (1-\varepsilon)\mathbb{E}(U) > 0 \\ 0, & \text{if } U - (1-\varepsilon)\mathbb{E}(U) \leq 0. \end{cases}$$

It can be seen that $\mathbb{E}(V') \geq \mathbb{E}(V)$ and $\mathbb{E}(V'^2) \leq \mathbb{E}(V^2)$. We then have that

$$\begin{aligned} \Pr(U > (1-\varepsilon)\mathbb{E}(U)) &= \Pr(V > 0) \\ &= \Pr(V' > 0) \\ &> \frac{(\mathbb{E}(V'))^2}{\mathbb{E}(V'^2)} \end{aligned} \quad (75)$$

$$\begin{aligned} &\geq \frac{(\mathbb{E}(V))^2}{\mathbb{E}(V^2)} \\ &\geq \frac{\varepsilon^2 n^{2\nu-2l_2 k_1 r}}{n^{\nu-l_2 k_1 r} - n^{\nu-2l_2 k_1 r} + \varepsilon^2 n^{2\nu-2l_2 k_1 r}} \end{aligned} \quad (76)$$

$$\begin{aligned} &\geq 1 - \frac{1}{\varepsilon^2} n^{-(\nu-l_2 k_1 r)} \\ &\geq 1 - \frac{1}{\varepsilon^2} n^{-(3l_1/2-l_2 k_1 r)}, \end{aligned} \quad (77)$$

where inequality (75) is due to the second-moment method. Inequality (76) follows by substituting the bound on $\mathbb{E}(V)$ and $\mathbb{E}(V^2)$ from equations (73) and (74) respectively, and inequality (77) holds since $\nu \geq 3l_1/2$. The equations above tell us that the actual number of combinations leading to a specific tuple $(w_1^{(\lambda L+1)}, w_1^{(\lambda L+2)}, \dots, w_1^{(L)})$ is greater than $(1-\varepsilon)$ times the expected number of combinations leading to it with high probability. By a similar argument that we omit here, we can also show that the actual number of combinations leading to $(w_1^{(\lambda L+1)}, w_1^{(\lambda L+2)}, \dots, w_1^{(L)})$ is less than $(1+\varepsilon)$ times the expected number of combinations with high probability. In conclusion, with probability at least $1 - \varepsilon^{-2} n^{-\mathcal{O}(\sqrt{n}/(\log n))}$ the variable U is concentrated in the range $(1 \pm \varepsilon)\mathbb{E}(U)$. We then take a union bound over all possible $(n^{l_2 k_1 r})$ tuples $(\mathbf{W}^{(\lambda L+1)}, \dots, \mathbf{W}^{(L)})$ of inner parity messages. This implies that with probability at least $1 - \varepsilon^{-2} n^{-\mathcal{O}(\sqrt{n}/(\log n))}$, every possible tuple of inner parity messages correspond to close to $n^{\nu-l_2 k_1 r}$ many inner systematic messages $(\mathbf{W}^{(1)}, \dots, \mathbf{W}^{(\lambda L)})$. Setting $\varepsilon = 1/n$ gives us the desired result. \square

With the help of Claim 11, we are able to show that the n -letter distribution p_1 on \vec{y}_w is pretty close to the corresponding “chunk-wise independent” product distribution when considering all the parity chunks.

Lemma 4 (Restated). *With probability at least $1 - \mathcal{O}(n^{-\delta}/\log n)$ over channel noise to Willie and concatenated code design, the n -letter distribution p_1 on \vec{y}_w is close to the “chunk-wise independent” product distribution, i.e.,*

$$\frac{1}{2} \sum_{\vec{y}_w^{(1)} \dots \vec{y}_w^{(L)}} \left| p_1^{(1)}(\vec{y}_w^{(1)}) \dots p_1^{(L)}(\vec{y}_w^{(L)}) - p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(L)}) \right| \leq \frac{1}{2} n^{-1}.$$

Proof: Recall that due to the independence of $\vec{y}_w^{(i)}$ on the systematic chunks, the n -letter distribution p_1 can be written as

$$\begin{aligned} p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(L)}) &= p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(\lambda L)}) \cdot p_1(\vec{y}_w^{(\lambda L+1)}, \dots, \vec{y}_w^{(L)} | \vec{y}_w^{(1)}, \dots, \vec{y}_w^{(\lambda L)}) \\ &= p_1^{(1)}(\vec{y}_w^{(1)}) \dots p_1^{(\lambda L)}(\vec{y}_w^{(\lambda L)}) \cdot p_1(\vec{y}_w^{(\lambda L+1)}, \dots, \vec{y}_w^{(L)} | \vec{y}_w^{(1)}, \dots, \vec{y}_w^{(\lambda L)}). \end{aligned} \quad (78)$$

Note that with probability (over channel noise to Willie and concatenated code design) at least $1 - \mathcal{O}(n^{-\delta}/\log n)$,

$$p_1\left(\vec{y}_w^{(\lambda L+1)}, \dots, \vec{y}_w^{(L)} | \vec{y}_w^{(1)}, \dots, \vec{y}_w^{(\lambda L)}\right) \quad (79)$$

$$= \sum_{w^{(\lambda L+1)}, \dots, w^{(L)}} p_1\left(\vec{y}_w^{(\lambda L+1)}, \dots, \vec{y}_w^{(L)} | \mathbf{W}^{(\lambda L+1)} = w^{(\lambda L+1)}, \dots, \mathbf{W}^{(L)} = w^{(L)}, \vec{y}_w^{(1)}, \dots, \vec{y}_w^{(\lambda L)}\right) \cdot p\left(\mathbf{W}^{(\lambda L+1)} = w^{(\lambda L+1)}, \dots, \mathbf{W}^{(L)} = w^{(L)} | \vec{y}_w^{(1)}, \dots, \vec{y}_w^{(\lambda L)}\right) \quad (80)$$

$$= \sum_{w^{(\lambda L+1)}, \dots, w^{(L)}} p_1\left(\vec{y}_w^{(\lambda L+1)}, \dots, \vec{y}_w^{(L)} | \mathbf{W}^{(\lambda L+1)} = w^{(\lambda L+1)}, \dots, \mathbf{W}^{(L)} = w^{(L)}\right) \frac{1}{|W^{(L)}|^{l_2}} (1 \pm n^{-1}) \quad (81)$$

$$= \sum_{w^{(\lambda L+1)}, \dots, w^{(L)}} p_1^{(\lambda L+1)}\left(\vec{y}_w^{(\lambda L+1)} | \mathbf{W}^{(\lambda L+1)} = w^{(\lambda L+1)}\right) \dots p_1^{(L)}\left(\vec{y}_w^{(L)} | \mathbf{W}^{(L)} = w^{(L)}\right) \cdot \frac{1}{|W^{(L)}|^{l_2}} (1 \pm n^{-1}) \quad (82)$$

$$= \sum_{w^{(\lambda L+1)}} p_1^{(\lambda L+1)}\left(\vec{y}_w^{(\lambda L+1)} | \mathbf{W}^{(\lambda L+1)} = w^{(\lambda L+1)}\right) \frac{1}{|W^{(\lambda L+1)}|} \dots \sum_{w^{(L)}} p_1^{(L)}\left(\vec{y}_w^{(L)} | \mathbf{W}^{(L)} = w^{(L)}\right) \frac{1}{|W^{(L)}|} \cdot (1 \pm n^{-1}) \quad (83)$$

$$= p_1^{(\lambda L+1)}\left(\vec{y}_w^{(\lambda L+1)}\right) \dots p_1^{(L)}\left(\vec{y}_w^{(L)}\right) \cdot (1 \pm n^{-1}). \quad (84)$$

We abuse notation in the above set of equations somewhat – the terms in (79) and (80) are mathematical expressions with a value, whereas those in (81)-(84) define intervals. For instance, in (81), the interval under consideration equals

$$\left[\frac{1}{|W^{(L)}|^{l_2}} (1 - n^{-1}), \frac{1}{|W^{(L)}|^{l_2}} (1 + n^{-1}) \right].$$

Hence the set of equations above should be understood as meaning that the expression in (79) lies in the interval in (84). Equation (80) follows from the total probability theorem. Equation (81) holds (with probability at least $1 - \mathcal{O}(n^{-\delta}/\log n)$) because firstly the received vectors of the systematic chunks $(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(\lambda L)})$, the inner-messages of the parity chunks $(\mathbf{W}^{(\lambda L+1)}, \dots, \mathbf{W}^{(L)})$, and the received vectors of the parity chunks $(\vec{y}_w^{(\lambda L+1)}, \dots, \vec{y}_w^{(L)})$, form a Markov chain, and secondly, because of Claim 11. Equation (82) is obtained since the channel $p(\vec{y}_w^{(i)} | \vec{x}^{(i)})$ is memoryless. Equation (83) follows by interchanging the order of summations in (82). Finally, combining (84) with (78) we have

$$\begin{aligned} p_1\left(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(L)}\right) &= p_1^{(1)}\left(\vec{y}_w^{(1)}\right) \dots p_1^{(\lambda L)}\left(\vec{y}_w^{(\lambda L)}\right) \cdot p_1\left(\vec{y}_w^{(\lambda L+1)}, \dots, \vec{y}_w^{(L)} | \vec{y}_w^{(1)}, \dots, \vec{y}_w^{(\lambda L)}\right) \\ &= p_1^{(1)}\left(\vec{y}_w^{(1)}\right) \dots p_1^{(L)}\left(\vec{y}_w^{(L)}\right) (1 \pm n^{-1}) \end{aligned}$$

The equations above mean that the n -letter distribution p_1 on \vec{y}_w is pretty close to the corresponding “chunk-wise independent” product distribution. Equipped with this fact, we conclude that

$$\begin{aligned} &\frac{1}{2} \sum_{\vec{y}_w^{(1)} \dots \vec{y}_w^{(L)}} \left| p_1^{(1)}(\vec{y}_w^{(1)}) \dots p_1^{(L)}(\vec{y}_w^{(L)}) - p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(L)}) \right| \\ &\leq \frac{1}{2} n^{-1} \sum_{\vec{y}_w^{(1)} \dots \vec{y}_w^{(L)}} p_1^{(1)}(\vec{y}_w^{(1)}) \dots p_1^{(L)}(\vec{y}_w^{(L)}) \\ &\leq \frac{1}{2} n^{-1}. \end{aligned}$$

□

By combining Lemmas 2, 3 and 4, we are able to show that with probability at least $1 - \mathcal{O}(n^{-\delta}/\log n)$, the variational distance between p_0 and p_1 is bounded from above as

$$\begin{aligned} \mathbb{V}(p_0, p_1) &\leq \mathbb{V}(p_0, \mathbb{E}_{\mathcal{C}}(p_1)) + \mathbb{V}(\mathbb{E}_{\mathcal{C}}(p_1), p_1) \\ &\leq \mathbb{V}(p_0, \mathbb{E}_{\mathcal{C}}(p_1)) + \frac{1}{2} \sum_{\vec{y}_w^{(1)} \dots \vec{y}_w^{(L)}} \left| p_1^{(1)}(\vec{y}_w^{(1)}) \dots p_1^{(L)}(\vec{y}_w^{(L)}) - p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(L)}) \right| \\ &\quad + \frac{1}{2} \sum_{\vec{y}_w^{(1)} \dots \vec{y}_w^{(L)}} \left| \mathbb{E}_{\mathcal{C}}(p_1(\vec{y}_w^{(1)}, \dots, \vec{y}_w^{(L)})) - p_1^{(1)}(\vec{y}_w^{(1)}) \dots p_1^{(L)}(\vec{y}_w^{(L)}) \right| \\ &\leq \epsilon_d + n^{-\delta/4} + \frac{1}{2} n^{-1} \\ &\leq \epsilon_d + 2n^{-\delta/4}, \end{aligned} \quad (85)$$

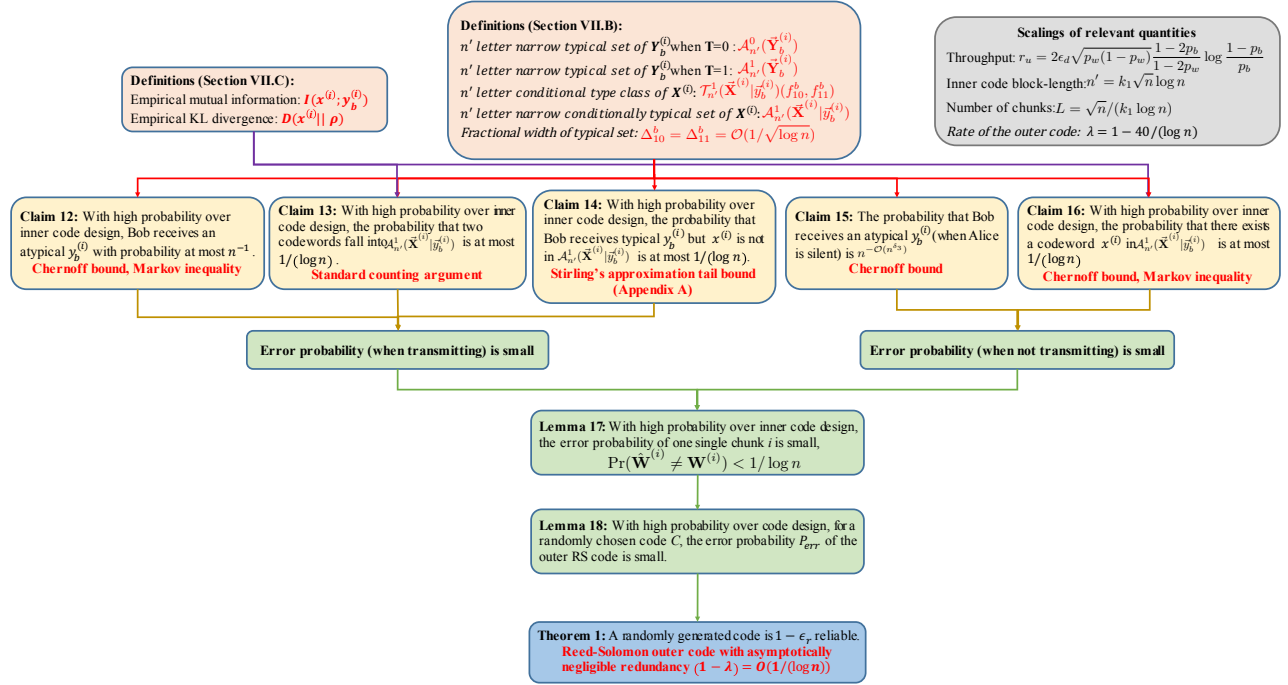


Fig. 5: A road-map of our proof that our codes are highly reliable with high probability.

This completes the proof of deniability of our proposed codes, as in Property 3) in Theorem 1.

IX. PROOF OF RELIABILITY

In this section we show that with high probability over the concatenated code ensemble \mathcal{C}^{cc} , the reliability of a randomly chosen code \mathcal{C} is at least $1 - \exp(-\sqrt{n}/(k_1(\log n)^2))$. Figure 5 is a flow-chart summarizing our proof of reliability. On receiving \vec{y}_b , Bob first partitions \vec{y}_b into L chunks $\vec{y}_b^{(1)}, \vec{y}_b^{(2)}, \dots, \vec{y}_b^{(L)}$. For each chunk i , Bob decodes the inner message $\hat{W}^{(i)} = \Gamma_{in}^{(i)}(\vec{y}_b^{(i)})$ by using the inner decoder $\Gamma_{in}^{(i)}(\cdot)$, and then reconstructs \hat{M} from Reed-Solomon code. We now elaborate on the **decoding rule** of Bob's inner decoder $\Gamma_{in}^{(i)}(\cdot)$, for reconstructing the inner message $\hat{W}^{(i)}$ as follows.

Decoding Rule $\hat{W}^{(i)}(\vec{y}_b^{(i)})$ for reconstructing the inner message on chunk i :

- 1) If $\vec{y}_b^{(i)} \in \mathcal{A}_{n'}^0(\vec{Y}_b^{(i)}) \setminus \mathcal{A}_{n'}^1(\vec{Y}_b^{(i)})$ and there is no codeword falling into the conditionally typical set $\mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_b^{(i)})$ with respect to $\vec{y}_b^{(i)}$, then Bob decodes $\hat{W}^{(i)} = 0$.
- 2) If $\vec{y}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{Y}_b^{(i)})$, then
 - a) If there is exactly one inner-message $w^{(i)}$ such that $\vec{x}^{(i)}(w^{(i)}) \in \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_b^{(i)})$, then Bob decodes $\hat{W}^{(i)} = w^{(i)}$.
 - b) If there exists two inner codewords $w^{(i)}$ and $w'^{(i)}$ such that $w^{(i)} \neq w'^{(i)}$, $\vec{x}^{(i)}(w^{(i)}) \in \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_b^{(i)})$ and $\vec{x}^{(i)}(w'^{(i)}) \in \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_b^{(i)})$, then Bob outputs an error.
 - c) If there does not exist a codeword $w^{(i)}$ such that $\vec{x}^{(i)}(w^{(i)}) \in \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_b^{(i)})$, then Bob decodes $\hat{W}^{(i)} = 0$.
- 3) If $\vec{y}_b^{(i)}$ is neither in $\mathcal{A}_{n'}^0(\vec{Y}_b^{(i)})$ nor in $\mathcal{A}_{n'}^1(\vec{Y}_b^{(i)})$, then Bob outputs an error.

In the following, we first consider the probability of decoding error of one single chunk, $\Pr(\hat{W}^{(i)} \neq W^{(i)})$, under the decoder $\Gamma_{in}^{(i)}(\cdot)$ in Claims 12-16 and Lemma 17, and then analyze the probability of error P_{err} of the outer RS code in Lemma 18. Figure 6 illustrates the potential error events of Bob's decoder $\Gamma(\cdot)$.

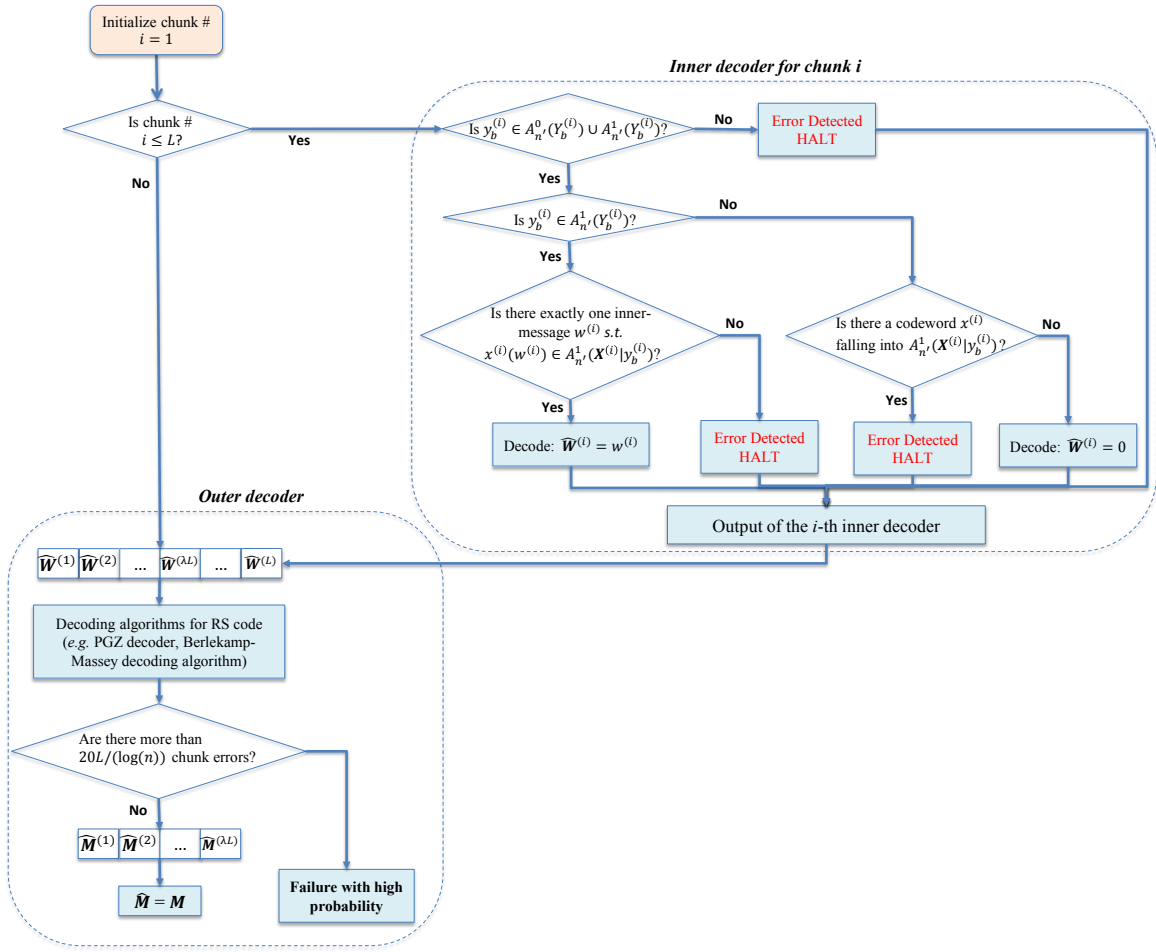


Fig. 6: A flow-chart describing potential error events.

A. Probability of decoding error of one single chunk

When Alice's transmission status $\mathbf{T} = 1$, without loss of generality, we assume the inner-message $\vec{x}^{(i)}(w^{(i)})$ is transmitted. Since the inner-messages (for the i -th chunk) are equiprobable and each codeword is generated i.i.d., the analysis of error probability is the same no matter which codeword is transmitted. Therefore, the probability of error (a) when Alice is transmitting is defined as

$$\Pr \left(\hat{\mathbf{W}}^{(i)} \neq w^{(i)} | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right).$$

When Alice's transmission status $\mathbf{T} = 0$, the probability of error (b) is defined as

$$\Pr \left(\hat{\mathbf{W}}^{(i)} \neq 0 | \mathbf{T} = 0 \right).$$

In the following, we show that both the probability of error when $\mathbf{T} = 0$ and the probability of error when $\mathbf{T} = 1$ go to 0 asymptotically.

Proof of (a): From Bob's decoding rule, the probability of error when transmitting ($\mathbf{T} = 1$) can be expanded as follows.

$$\begin{aligned} & \Pr_{\vec{\mathbf{Z}}_b} \left(\hat{\mathbf{W}}^{(i)} \neq w^{(i)} | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \\ & \leq \Pr_{\vec{\mathbf{Z}}_b} \left(\vec{\mathbf{Y}}_b^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \end{aligned} \quad (86)$$

$$+ \Pr_{\vec{\mathbf{Z}}_b, \vec{\mathbf{X}}^{(i)}} \left(\exists \vec{x}'^{(i)} \in \mathcal{C}^{(i)} \text{ s.t. } \vec{x}'^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)}), \vec{\mathbf{Y}}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \quad (87)$$

$$+ \Pr_{\vec{\mathbf{Z}}_b} \left(\vec{x}^{(i)}(w^{(i)}) \notin \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)}), \vec{\mathbf{Y}}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right). \quad (88)$$

The term in (86) corresponds to the probability of receiving an atypical $\vec{y}_b^{(i)}$. The term in (87) corresponds to the probability that Bob receives a typical $\vec{y}_b^{(i)}$, but there exists another codeword $\vec{x}'^{(i)} \neq \vec{x}^{(i)}$ falling into the conditionally typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)})$. The term in (88) corresponds to the probability that Bob receives a typical $\vec{y}_b^{(i)}$, but the true codeword $\vec{x}^{(i)}$ does not belong to the conditionally typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)})$.

In Claim 12-14, we present that the probability of the three error components, presented in (86), (87) and (88), goes to 0 asymptotically when n goes to infinity. In Claim 12, we set $\Delta_{*1}^{b,(1)} = n^{-1/4+\delta/2}$ (recall that $\Delta_{*1}^{b,(1)}$ is the parameter, defined in Section VII, specifying the “width” of the narrow typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)})$).

Claim 12 (Term in (86)). *With probability at least $1 - n^{-\frac{1}{3}k_1(\rho * p_b)n^\delta \log e + 1}$ over inner code design, the probability that Bob receives an atypical $\vec{y}_b^{(i)}$ is bounded from above as*

$$\Pr_{\vec{\mathbf{Z}}_b} \left(\vec{\mathbf{Y}}_b^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \leq n^{-1}.$$

Proof: The probability (over inner code design) that Bob receives an atypical $\vec{y}_b^{(i)}$ equals

$$\begin{aligned} \mathbb{E}_{\mathcal{C}^{(i)}} \left(\Pr_{\vec{\mathbf{Z}}_b} \left(\vec{\mathbf{Y}}_b^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \right) &= \Pr_{\vec{\mathbf{X}}^{(i)}, \vec{\mathbf{Z}}_b} \left(\vec{\mathbf{Y}}_b^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \\ &\leq 2 \exp \left(-\frac{1}{3} (\Delta_{*1}^{b,(1)})^2 \rho * p_b n' \right) \end{aligned} \quad (89)$$

$$= n^{-\frac{1}{3}k_1\rho * p_b n^\delta \log e}. \quad (90)$$

Equation (89) follows from the Chernoff bound, since the narrow typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)})$ is centered at $\rho * p_b$ and with width $\Delta_{*1}^{b,(1)}$. Equation (90) is obtained by substituting $\Delta_{*1}^{b,(1)}$ as $n^{-1/4+\delta/2}$. By applying Markov's inequality, we obtain

$$\Pr_{\mathcal{C}^{(i)}} \left(\Pr_{\vec{\mathbf{Z}}_b} \left(\vec{\mathbf{Y}}_b^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \geq n^{-1} \right) \leq n^{-\frac{1}{3}k_1\rho * p_b n^\delta \log e + 1},$$

which means with probability at least $1 - n^{-\frac{1}{3}k_1\rho * p_b n^\delta \log e + 1}$ over inner code design, the probability of receiving an atypical $\vec{y}_b^{(i)}$ is less than n^{-1} . This completes the proof of Claim 12. \square

Claim 13 (Term in (87)). *With probability at least $1 - \mathcal{O} \left((\log n)^2 \cdot 2^{-(\log n)^{2/3}} \right)$ over inner code design, the probability that Bob receives a typical $\vec{y}_b^{(i)}$ and there exists another codeword $\vec{x}'^{(i)} \neq \vec{x}^{(i)}$ falling into the conditionally typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)})$ is bounded from above as*

$$\Pr_{\vec{\mathbf{Z}}_b} \left(\exists \vec{x}'^{(i)} \in \mathcal{C}^{(i)} \text{ s.t. } \vec{x}'^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)}), \vec{\mathbf{Y}}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \leq 1/(\log n). \quad (91)$$

Proof: We first note that the expected number of inner codewords $\vec{x}'^{(i)}$ in chunk i falling into the typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)})$ equals the probability (over inner code design) of a single inner codeword falling into the typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)})$ times the size $|\mathcal{C}^{(i)}|$ of the inner codebook, and hence we have

$$\begin{aligned} & \mathbb{E}_{\mathcal{C}^{(i)}} \left[\Pr_{\vec{\mathbf{Z}}_b} \left(\exists \vec{x}'^{(i)} \in \mathcal{C}^{(i)} \text{ s.t. } \vec{x}'^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)}), \vec{\mathbf{Y}}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \right] \\ &= \Pr_{\vec{\mathbf{X}}^{(i)}, \vec{\mathbf{Z}}_b} \left(\vec{\mathbf{X}}^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)}), \vec{\mathbf{Y}}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \cdot |\mathcal{C}^{(i)}|. \end{aligned}$$

The size of the codebook $\mathcal{C}^{(i)}$ equals $2^{r' \log n}$, as defined in Section V. In the following, we will focus on the term

$$\Pr_{\vec{\mathbf{X}}^{(i)}, \vec{\mathbf{Z}}_b} \left(\vec{\mathbf{X}}^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)}), \vec{\mathbf{Y}}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right). \quad (92)$$

Note that the term in (92) is the probability of a single inner codeword falls into the conditionally typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)})$ averaged over all typical $\vec{y}_b^{(i)}$, and clearly it is less than this probability with respect to a typical $\vec{y}_b^{(i)}$ that maximizes it.

$$\Pr_{\vec{\mathbf{X}}^{(i)}, \vec{\mathbf{Z}}_b} \left(\vec{\mathbf{X}}^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)}), \vec{\mathbf{Y}}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \quad (93)$$

$$= \sum_{\vec{y}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)})} p(\vec{y}_b^{(i)}) \Pr_{\vec{\mathbf{X}}^{(i)}} \left(\vec{\mathbf{X}}^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \quad (94)$$

$$\leq \max_{\vec{y}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)})} \Pr_{\vec{\mathbf{X}}^{(i)}} \left(\vec{\mathbf{X}}^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right). \quad (95)$$

For any typical $\vec{y}_b^{(i)}$, the probability that a single inner codeword falls into the conditionally typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)})$ is bounded from above as

$$\begin{aligned} & \Pr_{\vec{\mathbf{X}}^{(i)}} \left(\vec{\mathbf{X}}^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \\ &= \sum_{\vec{x}^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)})} p(\vec{x}^{(i)}) \\ &= \sum_{f_{10}^b, f_{11}^b} \left(\sum_{\vec{x}^{(i)} \in \mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)})(f_{10}^b, f_{11}^b)} p(\vec{x}^{(i)}) \right) \end{aligned} \quad (96)$$

$$= \sum_{f_{10}^b, f_{11}^b} \binom{n' (f_{01}^b + f_{11}^b)}{n' f_{11}^b} \rho^{n' f_{11}^b} (1 - \rho)^{n' f_{01}^b} \binom{n' (f_{00}^b + f_{10}^b)}{n' f_{10}^b} \rho^{n' f_{10}^b} (1 - \rho)^{n' f_{00}^b} \quad (97)$$

$$\leq \sum_{f_{10}^b, f_{11}^b} \cdot 2^{-k_1 \sqrt{n} \log n [I(\vec{x}^{(i)}, \vec{y}_b^{(i)}) + D(\vec{x}^{(i)} || \rho)]} \quad (98)$$

$$\leq 2\rho p_b \frac{n'}{(\log n)^{1/2}} \cdot 2\rho(1 - p_b) \frac{n'}{(\log n)^{1/2}} \cdot 2^{-k_1 \sqrt{n} \log n [I(\vec{x}^{(i)}, \vec{y}_b^{(i)}) + D(\vec{x}^{(i)} || \rho)]} \quad (99)$$

$$= 4k_1^2 k_2^2 p_b (1 - p_b) (\log n) \cdot n^{-2k_1 \epsilon_d \sqrt{p_w(1-p_w)} \frac{1-2p_b}{1-2p_w} \log \left(\frac{1-p_b}{p_b} \right)}. \quad (100)$$

Equation (96) decomposes the conditionally typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)})$ into the conditional type classes $\mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)})$ that comprise it. To obtain equation (97), we use standard counting argument to calculate the probability that $\vec{\mathbf{X}}^{(i)}$ satisfies the condition of one type class $\mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)})(f_{10}^b, f_{11}^b)$ given a typical $\vec{y}_b^{(i)}$. Equation (98) follows from the Stirling's approximation, as well as the definition of empirical mutual information between $\vec{x}^{(i)}$ and $\vec{y}_b^{(i)}$, and the definition of empirical Kullback-Leibler divergence between $\vec{x}^{(i)}$ and ρ . Equation (99) follows since the number of (typical) conditional type class is bounded from above by¹⁴

$$2\rho p_b \frac{n'}{(\log n)^{1/2}} \cdot 2\rho(1 - p_b) \frac{n'}{(\log n)^{1/2}}.$$

In equation (100), we use the fact that

$$I(\vec{x}^{(i)}, \vec{y}_b^{(i)}) + D(\vec{x}^{(i)} || \rho) = \rho(1 - 2p_b) \log \left(\frac{1 - p_b}{p_b} \right) + o(1/\sqrt{n}) \quad (101)$$

$$= \frac{k_2}{\sqrt{n}} (1 - 2p_b) \log \left(\frac{1 - p_b}{p_b} \right) + o(1/\sqrt{n}) \quad (102)$$

$$= 2\epsilon_d \sqrt{\frac{p_w(1-p_w)}{n}} \frac{1 - 2p_b}{1 - 2p_w} \log \left(\frac{1 - p_b}{p_b} \right) + o(1/\sqrt{n}), \quad (103)$$

¹⁴Recall that we set the $\Delta_{10}^b = \Delta_{11}^b = 1/(\log n)^{1/2}$, which specify the “width” of the conditionally typical set.

where equation (101) is formally proved in Appendix D. Equations (102) and (103) follows from

$$\rho = k_2/\sqrt{n} = \frac{2\epsilon_d\sqrt{p_w(1-p_w)}}{(1-2p_w)\sqrt{n}},$$

where k_2 is first defined in equation (2), Section IV. Returning now to equation (93)-(95), we have

$$\begin{aligned} & \Pr_{\vec{\mathbf{x}}^{(i)}, \vec{\mathbf{z}}_b} \left(\vec{\mathbf{x}}'^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{x}}^{(i)} | \vec{\mathbf{y}}_b^{(i)}), \vec{\mathbf{y}}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{y}}_b^{(i)}) | \vec{\mathbf{x}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \\ & \leq \max_{\vec{y}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{y}}_b^{(i)})} \Pr_{\vec{\mathbf{x}}'^{(i)}} \left(\vec{\mathbf{x}}'^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{x}}^{(i)} | \vec{y}_b^{(i)}) | \vec{\mathbf{x}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \\ & \leq 4k_1^2 k_2^2 p_b (1-p_b) (\log n) \cdot n^{-2k_1\epsilon_d\sqrt{p_w(1-p_w)}\frac{1-2p_b}{1-2p_w}\log\left(\frac{1-p_b}{p_b}\right)}. \end{aligned}$$

Since the size of inner codebook is $2^{r' \log n}$, where

$$r' = rk_1/\lambda = r_u(1 - (\log n)^{-1/3})k_1/\lambda = \left(2k_1\epsilon_d\sqrt{p_w(1-p_w)}\frac{1-2p_b}{1-2p_w}\log\left(\frac{1-p_b}{p_b}\right) \right) \cdot (1 - (\log n)^{-1/3})/\lambda,$$

the probability (over inner code design) that there exist another codeword $\vec{x}''^{(i)} \neq \vec{x}^{(i)}$ falling into the conditionally typical set is bounded from above as

$$\begin{aligned} & \mathbb{E}_{\mathcal{C}^{(i)}} \left[\Pr_{\vec{\mathbf{z}}_b} \left(\exists \vec{x}''^{(i)} \in \mathcal{C}^{(i)} \text{ s.t. } \vec{x}''^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{x}}^{(i)} | \vec{\mathbf{y}}_b^{(i)}), \vec{\mathbf{y}}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{y}}_b^{(i)}) | \vec{\mathbf{x}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \right] \\ & \leq 2^{r' \log n} \cdot 4k_1^2 k_2^2 p_b (1-p_b) (\log n) \cdot n^{-2k_1\epsilon_d\sqrt{p_w(1-p_w)}\frac{1-2p_b}{1-2p_w}\log\left(\frac{1-p_b}{p_b}\right)} \\ & = 4k_1^2 k_2^2 p_b (1-p_b) (\log n) \cdot n^{r' - 2k_1\epsilon_d\sqrt{p_w(1-p_w)}\frac{1-2p_b}{1-2p_w}\log\left(\frac{1-p_b}{p_b}\right)} \quad (104) \\ & = \mathcal{O} \left((\log n) \cdot 2^{-(\log n)^{2/3}} \right). \quad (105) \end{aligned}$$

By applying Markov's inequality, we obtain that with probability (over inner code design) at least $1 - \mathcal{O} \left((\log n)^2 \cdot 2^{-(\log n)^{2/3}} \right)$,

$$\Pr_{\vec{\mathbf{z}}_b} \left(\exists \vec{x}''^{(i)} \in \mathcal{C}^{(i)} \text{ s.t. } \vec{x}''^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{x}}^{(i)} | \vec{\mathbf{y}}_b^{(i)}), \vec{\mathbf{y}}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{y}}_b^{(i)}) | \vec{\mathbf{x}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \leq 1/(\log n).$$

This completes the proof of Claim 13. \square

Claim 14 (Term in (88)). *With probability at least $1 - \sqrt{2/(\pi k_1 k_2 p_b)} \cdot (\log n) \cdot 2^{-k_1 k_2 p_b (\log n)^{1/2}}$ over inner code design, the probability that Bob receives a typical $\vec{y}_b^{(i)}$ if a conditionally atypical codeword $\vec{x}^{(i)}$ is transmitted is bounded from above as*

$$\Pr_{\vec{\mathbf{z}}_b} \left(\vec{x}^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{x}}^{(i)} | \vec{\mathbf{y}}_b^{(i)}), \vec{\mathbf{y}}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{y}}_b^{(i)}) | \vec{\mathbf{x}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \leq 1/(\log n).$$

Proof: Following the same arguments as in equation (93)-(95), we have

$$\begin{aligned} & \Pr_{\vec{\mathbf{x}}^{(i)}, \vec{\mathbf{z}}_b} \left(\vec{\mathbf{x}}^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{x}}^{(i)} | \vec{\mathbf{y}}_b^{(i)}), \vec{\mathbf{y}}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{y}}_b^{(i)}) | \vec{\mathbf{x}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \\ & \leq \max_{\vec{y}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{y}}_b^{(i)})} \Pr_{\vec{\mathbf{x}}^{(i)}} \left(\vec{\mathbf{x}}^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{x}}^{(i)} | \vec{y}_b^{(i)}) | \vec{\mathbf{x}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right). \end{aligned}$$

In the following we calculate the probability that the true codeword $\vec{x}^{(i)}$ does not belong to the conditionally typical set

$\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)})$ for any typical $\vec{y}_b^{(i)}$.

$$\begin{aligned} & \Pr_{\vec{\mathbf{X}}^{(i)}} \left(\vec{\mathbf{X}}^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \\ &= \Pr_{\vec{\mathbf{X}}^{(i)}} \left(f_{10}^b \notin [(1 - \Delta_{10}^b) \rho p_w, (1 + \Delta_{10}^b) \rho p_w] \cup f_{11}^b \notin [(1 - \Delta_{11}^b) \rho(1 - p_w), (1 + \Delta_{11}^b) \rho(1 - p_w)] \right) \end{aligned} \quad (106)$$

$$\leq \sum_{i_1=k_1 k_2 p_b (\log n) (1 + \Delta_{10}^b)}^{k_1 \sqrt{n} \log n} \binom{k_1 \sqrt{n} \log n}{i_1} \left(\frac{k_2 p_b}{\sqrt{n}} \right)^{i_1} \left(1 - \frac{k_2 p_b}{\sqrt{n}} \right)^{k_1 \sqrt{n} \log n - i_1} \quad (107)$$

$$+ \sum_{i_2=0}^{k_1 k_2 p_b (\log n) (1 - \Delta_{10}^b)} \binom{k_1 \sqrt{n} \log n}{i_2} \left(\frac{k_2 p_b}{\sqrt{n}} \right)^{i_2} \left(1 - \frac{k_2 p_b}{\sqrt{n}} \right)^{k_1 \sqrt{n} \log n - i_2} \quad (108)$$

$$+ \sum_{i_3=k_1 k_2 (1-p_b) (\log n) (1 + \Delta_{11}^b)}^{k_1 \sqrt{n} \log n} \binom{k_1 \sqrt{n} \log n}{i_3} \left(\frac{k_2 (1-p_b)}{\sqrt{n}} \right)^{i_3} \left(1 - \frac{k_2 (1-p_b)}{\sqrt{n}} \right)^{k_1 \sqrt{n} \log n - i_3} \quad (109)$$

$$+ \sum_{i_4=0}^{k_1 k_2 (1-p_b) (\log n) (1 - \Delta_{11}^b)} \binom{k_1 \sqrt{n} \log n}{i_4} \left(\frac{k_2 (1-p_b)}{\sqrt{n}} \right)^{i_4} \left(1 - \frac{k_2 (1-p_b)}{\sqrt{n}} \right)^{k_1 \sqrt{n} \log n - i_4} \quad (110)$$

$$\leq \sqrt{2/(\pi k_1 k_2 p_b)} \cdot 2^{-k_1 k_2 p_b (\log n)^{1/2}}. \quad (111)$$

Equation (106) follows since the conditionally typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)})$, defined in Section VII, is characterized by the “center” ρp_w and $\rho(1 - p_w)$, as well as the “width” Δ_{10}^b and Δ_{11}^b . Equations (107)-(110) follow from standard counting arguments, as presented in Section VIII, (57)-(60). To obtain (111), we first use Stirling’s approximation and then bound each of the four terms in (107)-(110) from above, which is elaborated in Appendix B (it works since Δ_{10}^b and Δ_{11}^b are chosen to be $1/(\log n)^{1/2}$). Then we obtain

$$\begin{aligned} & \mathbb{E}_{\mathcal{C}^{(i)}} \left(\Pr_{\vec{\mathbf{Z}}_b} \left(\vec{x}^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)}), \vec{\mathbf{Y}}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \right) \\ &= \Pr_{\vec{\mathbf{X}}^{(i)}, \vec{\mathbf{Z}}_b} \left(\vec{\mathbf{X}}^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)}), \vec{\mathbf{Y}}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \\ &\leq \Pr_{\vec{\mathbf{X}}^{(i)}} \left(\vec{\mathbf{X}}^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \\ &\leq \sqrt{2/(\pi k_1 k_2 p_b)} \cdot 2^{-k_1 k_2 p_b (\log n)^{1/2}}. \end{aligned}$$

By applying Markov’s inequality, we have

$$\Pr_{\vec{\mathbf{Z}}_b} \left(\vec{x}^{(i)} \notin \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)}), \vec{\mathbf{Y}}_b^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) \leq 1/(\log n),$$

with probability over inner code design at least $1 - \sqrt{2/(\pi k_1 k_2 p_b)} \cdot (\log n) \cdot 2^{-k_1 k_2 p_b (\log n)^{1/2}}$. This completes the proof of Claim 14. \square

Proof of (b): When Alice’s transmission status $\mathbf{T} = 0$, the probability of error can be decomposed as

$$\Pr_{\vec{\mathbf{Z}}_b} \left(\hat{\mathbf{W}}^{(i)} \neq 0 | \mathbf{T} = 0 \right) \leq \Pr_{\vec{\mathbf{Z}}_b} \left(\vec{\mathbf{Y}}_b^{(i)} \notin \mathcal{A}_{n'}^0(\vec{\mathbf{Y}}_b^{(i)}) | \mathbf{T} = 0 \right) \quad (112)$$

$$+ \Pr_{\vec{\mathbf{Z}}_b} \left(\exists \vec{x}^{(i)} \in \mathcal{C}^{(i)} \text{ s.t. } \vec{x}^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)}), \vec{\mathbf{Y}}_b^{(i)} \in \mathcal{A}_{n'}^0(\vec{\mathbf{Y}}_b^{(i)}) \setminus \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \mathbf{T} = 0 \right). \quad (113)$$

The term in (112) corresponds to the probability that Bob receives an atypical $\vec{y}_b^{(i)}$ (with respect to $\mathbf{T} = 0$). The term in (113) corresponds to the probability that Bob receives a typical $\vec{y}_b^{(i)}$ (with respect to $\mathbf{T} = 0$) but there exists a codeword $\vec{x}^{(i)}$ falling into the conditionally typical set $\mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)})$. In Claim 15 and 16, we show that the terms in (112) and (113) decrease to 0 asymptotically as n goes to infinity.

In Claim 15, we set $\Delta_{*1}^{b,(0)}$ to be $n^{-1/4+\delta/2}$ (recall that $\Delta_{*1}^{b,(0)}$ is the parameter, defined in Section VII, specifying the “width” of the narrow typical set $\mathcal{A}_{n'}^0(\vec{\mathbf{Y}}_b^{(i)})$).

Claim 15 (Term in (112)). When Alice's transmission status $\mathbf{T} = 0$, the probability that Bob receives an atypical $\vec{y}_b^{(i)}$ (with respect to $\mathbf{T} = 0$) is bounded from above as

$$\Pr_{\vec{z}_b} \left(\vec{Y}_b^{(i)} \notin \mathcal{A}_{n'}^0(\vec{Y}_b^{(i)}) | \mathbf{T} = 0 \right) \leq 2n^{-\frac{1}{3}(\log e)k_1 p_b n^\delta}.$$

Proof: The narrow typical set $\mathcal{A}_{n'}^0(\vec{Y}_b^{(i)})$ when $\mathbf{T} = 0$ is centered at p_b with width $\Delta_{*1}^{b,(0)} = n^{-1/4+\delta/2}$. We then calculate the probability of receiving an atypical $\vec{y}_b^{(i)}$ as follows,

$$\begin{aligned} \Pr_{\vec{z}_b} \left(\vec{Y}_b^{(i)} \notin \mathcal{A}_{n'}^0(\vec{Y}_b^{(i)}) | \mathbf{T} = 0 \right) &= \Pr_{\vec{z}_b} \left(f_{*1}^b \notin \left[p_b \left(1 \pm \Delta_{*1}^{b,(0)} \right) \right] | \mathbf{T} = 0 \right) \\ &\leq 2 \exp \left(-\frac{1}{3} (\Delta_{*1}^{b,(0)})^2 p_b n' \right) \end{aligned} \quad (114)$$

$$\begin{aligned} &= 2 \exp \left(-\frac{1}{3} k_1 p_b n^\delta \log n \right) \\ &= 2n^{-\frac{1}{3}(\log e)k_1 p_b n^\delta}, \end{aligned} \quad (115)$$

where (114) follows from the Chernoff bound, and (115) is obtained by substituting the value of n' as $k_1 \sqrt{n}(\log n)$, and the value of $\Delta_{*1}^{b,(0)}$ as $n^{-1/4+\delta/2}$. This completes the proof of Claim 15. \square

Claim 16 (Term in (113)). With probability at least $1 - \mathcal{O} \left((\log n)^2 \cdot 2^{-(\log n)^{2/3}} \right)$ over inner code design, the probability that Bob receives a typical $\vec{y}_b^{(i)}$ (with respect to $\mathbf{T} = 0$) as well as there exists a codeword falling into the conditionally typical set $\mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{Y}_b^{(i)})$ is bounded from above as

$$\Pr_{\vec{z}_b} \left(\exists \vec{x}^{(i)} \in \mathcal{C}^{(i)} \text{ s.t. } \vec{x}^{(i)} \in \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{Y}_b^{(i)}), \vec{Y}_b^{(i)} \in \mathcal{A}_{n'}^0(\vec{Y}_b^{(i)}) \setminus \mathcal{A}_{n'}^1(\vec{Y}_b^{(i)}) | \mathbf{T} = 0 \right) \leq 1/\log n.$$

Proof: The proof of Claim 16 is similar to the proof of Claim 13. Since the expected number of inner codewords $\vec{x}^{(i)}$ falling into the typical set equals the probability of a single inner codeword falling into the typical set times the size of $|\mathcal{C}^{(i)}|$ of the inner codebook, we obtain

$$\begin{aligned} &\mathbb{E}_{\mathcal{C}^{(i)}} \left[\Pr_{\vec{z}_b} \left(\exists \vec{x}^{(i)} \in \mathcal{C}^{(i)} \text{ s.t. } \vec{x}^{(i)} \in \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{Y}_b^{(i)}), \vec{Y}_b^{(i)} \in \mathcal{A}_{n'}^0(\vec{Y}_b^{(i)}) \setminus \mathcal{A}_{n'}^1(\vec{Y}_b^{(i)}) | \mathbf{T} = 0 \right) \right] \\ &= \Pr_{\vec{X}^{(i)}, \vec{z}_b} \left(\vec{X}^{(i)} \in \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{Y}_b^{(i)}), \vec{Y}_b^{(i)} \in \mathcal{A}_{n'}^0(\vec{Y}_b^{(i)}) \setminus \mathcal{A}_{n'}^1(\vec{Y}_b^{(i)}) | \mathbf{T} = 0 \right) \cdot |\mathcal{C}^{(i)}|. \end{aligned} \quad (116)$$

Next, we note that the probability of a single inner codeword falling into the typical set, appeared in (116), is averaged over all typical $\vec{y}_b^{(i)}$ with respect to $\mathbf{T} = 0$, and then we bound this probability from above as

$$\Pr_{\vec{X}^{(i)}, \vec{z}_b} \left(\vec{X}^{(i)} \in \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{Y}_b^{(i)}), \vec{Y}_b^{(i)} \in \mathcal{A}_{n'}^0(\vec{Y}_b^{(i)}) \setminus \mathcal{A}_{n'}^1(\vec{Y}_b^{(i)}) | \mathbf{T} = 0 \right) \quad (117)$$

$$= \sum_{\vec{y}_b^{(i)} \in \mathcal{A}_{n'}^0(\vec{Y}_b^{(i)}) \setminus \mathcal{A}_{n'}^1(\vec{Y}_b^{(i)})} p(\vec{y}_b^{(i)}) \Pr_{\vec{X}^{(i)}} \left(\vec{X}^{(i)} \in \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_b^{(i)}) | \mathbf{T} = 0 \right) \quad (118)$$

$$\leq \max_{\vec{y}_b^{(i)} \in \mathcal{A}_{n'}^0(\vec{Y}_b^{(i)}) \setminus \mathcal{A}_{n'}^1(\vec{Y}_b^{(i)})} \Pr_{\vec{X}^{(i)}} \left(\vec{X}^{(i)} \in \mathcal{A}_{n'}^1(\vec{X}^{(i)} | \vec{y}_b^{(i)}) | \mathbf{T} = 0 \right). \quad (119)$$

For any $\vec{y}_b^{(i)} \in \mathcal{A}_{n'}^0(\vec{\mathbf{Y}}_b^{(i)}) \setminus \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)})$, the probability that a single inner codeword falls into the conditionally typical set equals

$$\begin{aligned}
& \Pr_{\vec{\mathbf{X}}^{(i)}} \left(\vec{\mathbf{X}}^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)}) | \mathbf{T} = 0 \right) \\
&= \sum_{\vec{x}^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)})} p(\vec{x}^{(i)}) \\
&= \sum_{f_{10}^b, f_{11}^b} \left(\sum_{\vec{x}^{(i)} \in \mathcal{T}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{y}_b^{(i)}) (f_{10}^b, f_{11}^b)} p(\vec{x}^{(i)}) \right) \\
&= \sum_{f_{10}^b, f_{11}^b} \binom{n' (f_{01}^b + f_{11}^b)}{n' f_{11}^b} \rho^{n' f_{11}^b} (1 - \rho)^{n' f_{01}^b} \binom{n' (f_{00}^b + f_{10}^b)}{n' f_{10}^b} \rho^{n' f_{10}^b} (1 - \rho)^{n' f_{00}^b} \\
&\leq 4k_1^2 k_2^2 p_b (1 - p_b) (\log n) \cdot n^{-2k_1 \epsilon_d \sqrt{p_w(1-p_w)} \frac{1-2p_b}{1-2p_w} \log \left(\frac{1-p_b}{p_b} \right)}.
\end{aligned}$$

Returning to equation (117), it then follows that the probability of a single inner codeword falling into the typical set is bounded from above as

$$\Pr_{\vec{\mathbf{X}}^{(i)}, \vec{\mathbf{Z}}_b} \left(\vec{\mathbf{X}}^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)}), \vec{\mathbf{Y}}_b^{(i)} \in \mathcal{A}_{n'}^0(\vec{\mathbf{Y}}_b^{(i)}) \setminus \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \mathbf{T} = 0 \right) \leq 4k_1^2 k_2^2 p_b (1 - p_b) (\log n) \cdot n^{-2k_1 \epsilon_d \sqrt{p_w(1-p_w)} \frac{1-2p_b}{1-2p_w} \log \left(\frac{1-p_b}{p_b} \right)},$$

and the probability (over inner code design) that there exists an inner codeword falling into the typical set is bounded from above as

$$\begin{aligned}
& \mathbb{E}_{\mathcal{C}^{(i)}} \left[\Pr_{\vec{\mathbf{Z}}_b} \left(\exists \vec{x}^{(i)} \in \mathcal{C}^{(i)} \text{ s.t. } \vec{x}^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)}), \vec{\mathbf{Y}}_b^{(i)} \in \mathcal{A}_{n'}^0(\vec{\mathbf{Y}}_b^{(i)}) \setminus \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \mathbf{T} = 0 \right) \right] \\
&= \Pr_{\vec{\mathbf{X}}^{(i)}, \vec{\mathbf{Z}}_b} \left(\vec{\mathbf{X}}^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)}), \vec{\mathbf{Y}}_b^{(i)} \in \mathcal{A}_{n'}^0(\vec{\mathbf{Y}}_b^{(i)}) \setminus \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \mathbf{T} = 0 \right) \cdot |\mathcal{C}^{(i)}| \\
&\leq 4k_1^2 k_2^2 p_b (1 - p_b) (\log n) \cdot n^{-2k_1 \epsilon_d \sqrt{p_w(1-p_w)} \frac{1-2p_b}{1-2p_w} \log \left(\frac{1-p_b}{p_b} \right)} \cdot 2^{r' \log n} \\
&\leq \mathcal{O} \left((\log n) \cdot 2^{-(\log n)^{2/3}} \right),
\end{aligned}$$

where $r' = rk_1/\lambda$. Finally, by the Markov inequality, we obtain that with probability (over inner code design) at least $1 - \mathcal{O} \left((\log n)^2 \cdot 2^{-(\log n)^{2/3}} \right)$,

$$\Pr_{\vec{\mathbf{Z}}_b} \left(\exists \vec{x}^{(i)} \in \mathcal{C}^{(i)} \text{ s.t. } \vec{x}^{(i)} \in \mathcal{A}_{n'}^1(\vec{\mathbf{X}}^{(i)} | \vec{\mathbf{Y}}_b^{(i)}), \vec{\mathbf{Y}}_b^{(i)} \in \mathcal{A}_{n'}^0(\vec{\mathbf{Y}}_b^{(i)}) \setminus \mathcal{A}_{n'}^1(\vec{\mathbf{Y}}_b^{(i)}) | \mathbf{T} = 0 \right) \leq 1/\log n.$$

This completes the proof of Claim 16. \square

Having proved Claims 12-16, it turns out that the probability of decoding error of one single chunk follows directly. For notational convenience we define $\zeta_{prob} = \sqrt{2/(\pi k_1 k_2 p_b)} (\log n) 2^{-k_1 k_2 p_b (\log n)^{1/2}}$, and then we have the following lemma.

Lemma 17. *With probability at least $1 - 4\zeta_{prob}$ over inner code design, the probability of error of the i -th chunk ($1 \leq i \leq L$) is bounded from above as*

$$\Pr \left(\hat{\mathbf{W}}^{(i)} \neq \mathbf{W}^{(i)} \right) < 6/\log n.$$

Proof: By the union bound and the fact that $(\log n)^2 \cdot 2^{-(\log n)^{2/3}}$ is $\mathcal{O} \left((\log n) \cdot 2^{-(\log n)^{1/2}} \right)$, we obtain that for sufficiently large n , the probability of error of one single chunk is bounded from above as

$$\begin{aligned}
\Pr \left(\hat{\mathbf{W}}^{(i)} \neq \mathbf{W}^{(i)} \right) &\leq \Pr_{\vec{\mathbf{Z}}_b} \left(\hat{\mathbf{W}}^{(i)} \neq w^{(i)} | \vec{\mathbf{X}}^{(i)} = \vec{x}^{(i)}(w^{(i)}), \mathbf{T} = 1 \right) + \Pr_{\vec{\mathbf{Z}}_b} \left(\hat{\mathbf{W}}^{(i)} \neq 0 | \mathbf{T} = 0 \right) \\
&\leq 3/\log n + n^{-1} + 2n^{-\frac{1}{3}(\log e)k_1 p_b n^{\delta}} \\
&\leq 6/\log n,
\end{aligned} \tag{120}$$

with probability (over inner code design) at least $1 - 4\zeta_{prob}$. Inequality (120) basically follows from Claims 12-16. This completes the proof of Lemma 17. \square

B. Probability of error of the outer RS code

Lemma 18. *With probability at least $1 - \exp(-\frac{4L}{3}\zeta_{prob})$ over concatenated code design, for a randomly chosen code \mathcal{C} , the probability of error P_{err} of the outer Reed-Solomon code is bounded from above as*

$$P_{err} \leq \exp(-2\sqrt{n}/(k_1(\log n)^2)).$$

Proof: Lemma 17 shows that with probability at least $1 - 4\zeta_{prob}$ over inner code design, the probability of error of a randomly chosen inner code $\Pr(\hat{\mathbf{W}}^{(i)} \neq \mathbf{W}^{(i)}) < 6/\log n$. An inner code (for chunk i) is said to be a *good inner code (for chunk i)* if the probability of error over the channel noise $p(\hat{y}_w^{(i)}|\mathbf{W}^{(i)})$ is bounded from above by $6/\log n$, and is said to be a *bad inner code (for chunk i)* otherwise. Let Λ_1 and Λ_2 be the number of chunk errors induced by good and bad inner codes respectively, that the RS outer code will need to correct. In the following we focus on the impact of good and bad inner codes on number of chunk in error.

(i) Impact of good inner codes on number of chunk in error: Since the number of good inner codes is at most L , and as shown in Lemma 17 the probability of error of good inner codes is bounded from above by $6/\log n$, it then follows that the expected number of chunk in error induced by good inner codes, $\mathbb{E}(\Lambda_1)$, is bounded from above by $6L/\log n$. Since the inner codes are chosen according to an i.i.d. distribution, by the Chernoff bound, with probability at least $1 - \exp(-2L/(\log n))$ over code design, the number of chunk in error induced by good inner codes is bounded from above by $12L/(\log n)$.

(ii) Impact of bad inner codes on number of chunk in error: As shown in Lemma 17, the probability of generating a bad inner code equals $4\zeta_{prob}$, and hence the expected number of bad inner codes equals $4L\zeta_{prob}$. Since the inner codes are chosen according to an i.i.d. distribution, by the Chernoff bound, with probability at least $1 - \exp(-\frac{4L}{3}\zeta_{prob})$ over code design, the number of bad inner codes is bounded from above by $8L\zeta_{prob}$, and hence the number of chunk in error induced by bad inner codes, Λ_2 , is bounded from above by $8L\zeta_{prob}$.

(iii) Concentration of overall inner codes in error: A concatenated code \mathcal{C} is said to be a *decent* code if the number of bad inner codes of \mathcal{C} is no more than $8L\zeta_{prob}$. From (ii) we know that with probability at least $1 - \exp(-\frac{4L}{3}\zeta_{prob})$ over code design, a randomly chosen code \mathcal{C} from the concatenated code ensemble $p(\mathcal{C}^{cc})$ is decent. Conditioned on the event that a decent code \mathcal{C} is chosen, it then follows from (i) that with probability at least $(1 - \exp(-2L/(\log n)))$, the number of chunk in error induced by good inner codes is bounded from above by $12L/(\log n)$, and hence the number of overall inner codes in error is bounded from above as

$$\Lambda_1 + \Lambda_2 \leq 12L/(\log n) + 8L\zeta_{prob} \leq 20L/(\log n).$$

Our outer Reed-Solomon code is able to correct $20L/(\log n)$ errors, since the number of parity chunks is $40L/(\log n)$. Therefore, with probability at least $1 - \exp(-\frac{4L}{3}\zeta_{prob})$ over concatenated code design, for a randomly chosen code \mathcal{C} , the probability of error P_{err} of the outer Reed-Solomon code is

$$\begin{aligned} P_{err} &\leq \Pr((\Lambda_1 + \Lambda_2) > 20L/(\log n)) \\ &\leq \exp(-2L/(\log n)) \\ &= \exp(-2\sqrt{n}/(k_1(\log n)^2)). \end{aligned}$$

This completes the proof of Lemma 18, as well as the proof of deniability of our proposed codes, as in Property 2) in Theorem 1. \square

ACKNOWLEDGEMENT

The authors would like to thank Andrej Bogdanov, Tongxin Li and Pak Hou Che for their valuable suggestions.

APPENDIX A

The Chernoff bound [35] is widely used in this work. Since there are many different versions of the Chernoff bound in the literature, and each version has a slightly different formulation, in this Appendix we explicitly state the version of the Chernoff bound [35] used throughout this work.

Suppose Q_1, \dots, Q_n are independent (but not necessarily identically distributed) random variables taking values in $\{0, 1\}$. We define Q as $Q_1 + \dots + Q_n$, and denote the expectation of Q by $\mathbb{E}(Q)$. Then for any $0 < \epsilon < 1$,

$$\begin{aligned} \Pr(Q \geq (1 + \epsilon)\mathbb{E}(Q)) &\leq e^{-\frac{\epsilon^2 \mathbb{E}(Q)}{3}}, \\ \Pr(Q \leq (1 - \epsilon)\mathbb{E}(Q)) &\leq e^{-\frac{\epsilon^2 \mathbb{E}(Q)}{2}} \leq e^{-\frac{\epsilon^2 \mathbb{E}(Q)}{3}}. \end{aligned}$$

APPENDIX B

Let $i_0 = k_1 k_2 p_w (\log n) (1 + \Delta_{10}^w)$ and an auxiliary function $h(i)$ be defined as

$$h(i) = \binom{k_1 \sqrt{n} \log n}{i} \left(\frac{k_2 p_w}{\sqrt{n}} \right)^i \left(1 - \frac{k_2 p_w}{\sqrt{n}} \right)^{k_1 \sqrt{n} \log n - i}.$$

Then, via Stirling's approximation [36, pp. 50-53], as n grows without bound, we can bound $h(i_0)$ from above as

$$\begin{aligned} h(i_0) &= \binom{k_1 \sqrt{n} \log n}{i_0} \left(\frac{k_2 p_w}{\sqrt{n}} \right)^{i_0} \left(1 - \frac{k_2 p_w}{\sqrt{n}} \right)^{k_1 \sqrt{n} \log n - i_0} \\ &\leq \frac{1}{\sqrt{2\pi i_0}} 2^{i_0 \log \left(\frac{e k_1 \sqrt{n} \log n}{i_0} \right)} 2^{i_0 \log \left(\frac{k_2 p_w}{\sqrt{n}} \right)} 2^{(k_1 \sqrt{n} \log n - i_0) \log \left(\frac{\sqrt{n} - k_2 p_w}{\sqrt{n}} \right)} \\ &= \frac{1}{\sqrt{2\pi i_0}} 2^{i_0 \log \left(\frac{e \sqrt{n}}{k_2 p_w (1 + \Delta_{10}^w)} \right)} 2^{i_0 \log \left(\frac{k_2 p_w}{\sqrt{n}} \right)} 2^{(k_1 \sqrt{n} \log n - i_0) \log \left(\frac{\sqrt{n} - k_2 p_w}{\sqrt{n}} \right)} \\ &= \frac{1}{\sqrt{2\pi i_0}} n^{k_1 k_2 p_w (1 + \Delta_{10}^w) \log \left(\frac{e}{1 + \Delta_{10}^w} \right) + [\sqrt{n} - k_2 p_w (1 + \Delta_{10}^w)] k_1 \log \left(\frac{\sqrt{n} - k_2 p_w}{\sqrt{n}} \right)} \\ &\stackrel{(a)}{=} \frac{1}{\sqrt{2\pi i_0}} n^{k_1 k_2 p_w (1 + \Delta_{10}^w) \log \left(\frac{e}{1 + \Delta_{10}^w} \right) + \frac{k_1 k_2 p_w \sqrt{n} \log e}{\sqrt{n} - k_2 p_w} + \mathcal{O}(n^{-1/2})} \end{aligned} \quad (121)$$

$$= \frac{1}{\sqrt{2\pi k_1 k_2 p_w (\log n) (1 + \Delta_{10}^w)}} n^{k_1 k_2 p_w \left((1 + \Delta_{10}^w) \log \left(\frac{e}{1 + \Delta_{10}^w} \right) - \log e \right)}, \quad (122)$$

where equality (121) follows from $\log \left(\frac{\sqrt{n} - k_2 p_w}{\sqrt{n}} \right) = -\frac{k_2 p_w}{\sqrt{n} - k_2 p_w} + \mathcal{O}(n^{-1})$, by applying Taylor's series expansion. We observe that the ratio between two successive terms is

$$\frac{h(i+1)}{h(i)} = \frac{k_1 \sqrt{n} \log n - i}{i+1} \cdot \frac{k_2 p_w}{\sqrt{n} - k_2 p_w}.$$

Hence for $i \geq i_0 = k_1 k_2 p_w (\log n) (1 + \Delta_{10}^w)$, we have

$$\frac{h(i+1)}{h(i)} \leq \frac{k_1 \sqrt{n} \log n - k_1 k_2 p_w (\log n) (1 + \Delta_{10}^w)}{k_1 k_2 p_w (\log n) (1 + \Delta_{10}^w) + 1} \cdot \frac{k_2 p_w}{\sqrt{n} - k_2 p_w} \leq \frac{1}{1 + \Delta_{10}^w}$$

This implies that the tail of the series $\{h(i)\}$ can be bounded from above by a geometric series as follows:

$$\begin{aligned} \sum_{i=i_0}^{k_1 \sqrt{n} \log n} h(i) &\leq h(i_0) \left[1 + \left(\frac{1}{1 + \Delta_{10}^w} \right) + \cdots + \left(\frac{1}{1 + \Delta_{10}^w} \right)^{k_1 \sqrt{n} \log n - i_0} \right] \\ &\leq h(i_0) \left[\sum_{j=0}^{\infty} \left(\frac{1}{1 + \Delta_{10}^w} \right)^j \right] \\ &= h(i_0) \left(\frac{1 + \Delta_{10}^w}{\Delta_{10}^w} \right). \end{aligned}$$

Substituting in the bound on $h(i_0)$ from Equation (122) gives us

$$\sum_{i=i_0}^{k_1 \sqrt{n} \log n} h(i) \leq \sqrt{\frac{1 + \Delta_{10}^w}{2\pi k_1 k_2 p_w (\Delta_{10}^w)^2 (\log n)}} \cdot n^{k_1 k_2 p_w \left((1 + \Delta_{10}^w) \log \left(\frac{e}{1 + \Delta_{10}^w} \right) - \log e \right)} \leq n^{-k_1 k_2 p_w f(\Delta_{10}^w)},$$

hence proving the term (61) in Section VIII-A. Similarly, one can also prove the term (62) in Section VIII-A, i.e.,

$$\sum_{i=k_1 k_2 (1-p_w) (\log n) (1 + \Delta_{11}^w)}^{k_1 \sqrt{n} \log n} \binom{k_1 \sqrt{n} \log n}{i} \left(\frac{k_2 (1-p_w)}{\sqrt{n}} \right)^i \left(1 - \frac{k_2 (1-p_w)}{\sqrt{n}} \right)^{k_1 \sqrt{n} \log n - i} \leq n^{-k_1 k_2 (1-p_w) f(\Delta_{11}^w)}.$$

APPENDIX C

We first suppose the generator matrix $G_{\lambda L \times L}$ of a general Reed-Solomon code has the form

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ \mu_1 & \mu_2 & \mu_3 & \dots & \mu_L \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_1^{\lambda L-1} & \mu_2^{\lambda L-1} & \mu_3^{\lambda L-1} & \dots & \mu_L^{\lambda L-1} \end{bmatrix}, \quad (123)$$

where $\mu_1, \mu_2, \dots, \mu_L$ are all distinct. The message $\vec{m} = [m_1, m_2, \dots, m_{\lambda L}]$ of the Reed-Solomon code is uniformly distributed over $\mathbb{F}_q^{\lambda L}$, and the codeword $\vec{w} = \vec{m} \cdot G_{\lambda L \times L} = [w_1, w_2, \dots, w_L] \in \mathbb{F}_q^L$, where $q = 2^{r' \log n}$ as specified in Section V. And the code is denoted by $\mathcal{C}_{RS} = \{\vec{w} : \vec{w} = \vec{m} \cdot G_{\lambda L \times L}, \forall \vec{m} \in \mathbb{F}_q^{\lambda L}\}$.

As a systematic Reed-Solomon code serves as our outer code, we can obtain the generator matrix $G'_{\lambda L \times L}$ of the systematic Reed-Solomon code by performing Gaussian elimination, i.e., $G'_{\lambda L \times L} = B^{-1} \cdot G_{\lambda L \times L} = \begin{bmatrix} I_{\lambda L \times \lambda L} & P \end{bmatrix}$, where B^{-1} is an invertible matrix and $I_{\lambda L \times \lambda L}$ is an identity matrix. Let's denote the systematic Reed-Solomon code by

$$\begin{aligned} \mathcal{C}_{SRS} &= \{\vec{w}' : \vec{w}' = \vec{m} \cdot G'_{\lambda L \times L}, \forall \vec{m} \in \mathbb{F}_q^{\lambda L}\} \\ &= \{\vec{w}' : \vec{w}' = \vec{m} \cdot B^{-1} G_{\lambda L \times L}, \forall \vec{m} \in \mathbb{F}_q^{\lambda L}\} \\ &= \{\vec{w}' : \vec{w}' = \vec{m} \cdot G_{\lambda L \times L}, \forall \vec{m} \in \mathbb{F}_q^{\lambda L}\}, \end{aligned} \quad (124)$$

where equation (124) holds since the linear mapping B^{-1} is bijective. The code \mathcal{C}_{SRS} with a systematic generator matrix $G'_{\lambda L \times L}$ is essentially same as \mathcal{C}_{RS} .

For the j -th location of \vec{w}' , $w'_j = [m_1 \ m_2 \ \dots \ m_{\lambda L}] \cdot [1 \ \mu_j^1 \ \dots \ \mu_j^{\lambda L}]^T = \sum_{i=0}^{\lambda L-1} m_{i+1} \mu_j^i$. For a specific parity inner-message vector $(w^{(\lambda L+1)}, \dots, w^{(L)})$ defined in Section VIII, the systematic inner-message that could cause it satisfies

$$\begin{cases} m_1 + m_2 \cdot \mu_{\lambda L+1} + m_3 \cdot \mu_{\lambda L+1}^2 + \dots + m_{\lambda L} \cdot \mu_{\lambda L+1}^{\lambda L-1} = w^{(\lambda L+1)}, \\ m_1 + m_2 \cdot \mu_{\lambda L+2} + m_3 \cdot \mu_{\lambda L+2}^2 + \dots + m_{\lambda L} \cdot \mu_{\lambda L+2}^{\lambda L-1} = w^{(\lambda L+2)}, \\ \vdots \\ m_1 + m_2 \cdot \mu_L + m_3 \cdot \mu_L^2 + \dots + m_{\lambda L} \cdot \mu_L^{\lambda L-1} = w^{(L)}. \end{cases}$$

Therefore, for any parity inner-message vector $(w^{(\lambda L+1)}, \dots, w^{(L)})$, the number of systematic inner-messages that could cause it equals $q^{\lambda L - (1-\lambda)L}$, since the null-space of the Vandermonde matrix

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ \mu_{\lambda L+1} & \mu_{\lambda L+2} & \dots & \mu_L \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{\lambda L+1}^{\lambda L-1} & \mu_{\lambda L+1}^{\lambda L-1} & \dots & \mu_L^{\lambda L-1} \end{bmatrix}$$

is $\lambda L - (1-\lambda)L$ -dimensional. Therefore we conclude the number of occurrences of any parity inner-message vector in the code \mathcal{C}_{SRS} are the same.

APPENDIX D

We aim to calculate the value of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ when $f_{*1}^w \in \rho * p_w(1 \pm \Delta_{*1}^w)$, $f_{10}^w \in \rho p_w(1 \pm \Delta_{10}^w)$, and $f_{11}^w \in \rho * (1 - p_w)(1 \pm \Delta_{11}^w)$. By definition, The first term $I(\vec{x}^{(i)}; \vec{y}_w^{(i)})$ can be expressed as

$$\begin{aligned} I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) &= \sum_{(j,j') \in \{0,1\} \times \{0,1\}} f_{jj'}^w \log \frac{f_{jj'}^w}{f_{j*}^w \cdot f_{*j'}^w} \\ &= f_{00}^w \log \left(\frac{f_{00}^w}{(1 - f_{1*}^w)(1 - f_{*1}^w)} \right) + f_{01}^w \log \left(\frac{f_{01}^w}{(1 - f_{1*}^w)f_{*1}^w} \right) + f_{10}^w \log \left(\frac{f_{10}^w}{f_{1*}^w(1 - f_{*1}^w)} \right) + f_{11}^w \log \left(\frac{f_{11}^w}{f_{1*}^w f_{*1}^w} \right) \\ &= (1 - f_{*1}^w - f_{10}^w) \log \left(\frac{1 - f_{*1}^w - f_{10}^w}{(1 - f_{10}^w - f_{11}^w)(1 - f_{*1}^w)} \right) + (f_{*1}^w - f_{11}^w) \log \left(\frac{f_{*1}^w - f_{11}^w}{(1 - f_{10}^w - f_{11}^w)f_{*1}^w} \right) \\ &\quad + f_{10}^w \log \left(\frac{f_{10}^w}{(f_{10}^w + f_{11}^w)(1 - f_{*1}^w)} \right) + f_{11}^w \log \left(\frac{f_{11}^w}{(f_{10}^w + f_{11}^w)f_{*1}^w} \right). \end{aligned}$$

It then follows that the partial derivative of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)})$ with respect to f_{*1}^w equals

$$\begin{aligned} \frac{\partial I(\vec{x}^{(i)}; \vec{y}_w^{(i)})}{\partial f_{*1}^w} &= -\log \left(\frac{1 - f_{*1}^w - f_{10}^w}{(1 - f_{10}^w - f_{11}^w)(1 - f_{*1}^w)} \right) - \frac{f_{10}^w}{1 - f_{*1}^w} + \log \left(\frac{f_{*1}^w - f_{11}^w}{(1 - f_{10}^w - f_{11}^w)f_{*1}^w} \right) + \frac{f_{11}^w}{f_{*1}^w} + \frac{f_{10}^w}{1 - f_{*1}^w} - \frac{f_{11}^w}{f_{*1}^w} \\ &= \log \left(\frac{f_{*1}^w - f_{11}^w}{(1 - f_{10}^w - f_{11}^w)f_{*1}^w} \cdot \frac{(1 - f_{10}^w - f_{11}^w)(1 - f_{*1}^w)}{1 - f_{*1}^w - f_{10}^w} \right) \\ &= \log \left(\frac{(f_{*1}^w - f_{11}^w)(1 - f_{*1}^w)}{f_{*1}^w(1 - f_{*1}^w - f_{10}^w)} \right). \end{aligned} \quad (125)$$

Similarly, the partial derivative of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)})$ with respect to f_{10}^w equals

$$\frac{\partial I(\vec{x}^{(i)}; \vec{y}_w^{(i)})}{\partial f_{10}^w} = \log \left(\frac{f_{10}^w(1 - f_{10}^w - f_{11}^w)}{(f_{10}^w + f_{11}^w)(1 - f_{*1}^w - f_{10}^w)} \right), \quad (126)$$

and the partial derivative of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)})$ with respect to f_{11}^w equals

$$\frac{\partial I(\vec{x}^{(i)}; \vec{y}_w^{(i)})}{\partial f_{11}^w} = \log \left(\frac{f_{11}^w(1 - f_{10}^w - f_{11}^w)}{(f_{10}^w + f_{11}^w)(f_{*1}^w - f_{11}^w)} \right). \quad (127)$$

The second term $D(\vec{x}^{(i)} \parallel \rho)$ can be expressed as

$$\begin{aligned} D(\vec{x}^{(i)} \parallel \rho) &= f_{0*} \log \frac{f_{0*}}{1 - \rho} + f_{1*} \log \frac{f_{1*}}{\rho} \\ &= (1 - f_{10}^w - f_{11}^w) \log \frac{(1 - f_{10}^w - f_{11}^w)}{1 - \rho} + (f_{10}^w + f_{11}^w) \log \frac{(f_{10}^w + f_{11}^w)}{\rho}. \end{aligned}$$

The partial derivative of $D(\vec{x}^{(i)} \parallel \rho)$ with respect to f_{10}^w equals

$$\frac{\partial D(\vec{x}^{(i)} \parallel \rho)}{\partial f_{10}^w} = \log \left(\frac{(1 - \rho)(f_{10}^w + f_{11}^w)}{\rho(1 - f_{10}^w - f_{11}^w)} \right),$$

and the partial derivative of $D(\vec{x}^{(i)} \parallel \rho)$ with respect to f_{11}^w also equals

$$\frac{\partial D(\vec{x}^{(i)} \parallel \rho)}{\partial f_{11}^w} = \log \left(\frac{(1 - \rho)(f_{10}^w + f_{11}^w)}{\rho(1 - f_{10}^w - f_{11}^w)} \right).$$

Therefore, the partial derivative of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ with respect to f_{10}^w is given as

$$\begin{aligned} \frac{\partial [I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)]}{\partial f_{10}^w} &= \log \left(\frac{f_{10}^w(1 - f_{10}^w - f_{11}^w)}{(f_{10}^w + f_{11}^w)(1 - f_{*1}^w - f_{10}^w)} \right) + \log \left(\frac{(1 - \rho)(f_{10}^w + f_{11}^w)}{\rho(1 - f_{10}^w - f_{11}^w)} \right) \\ &= \log \left(\frac{f_{10}^w(1 - \rho)}{\rho(1 - f_{*1}^w - f_{10}^w)} \right). \end{aligned} \quad (128)$$

Note that the value of term (128) is negative when $f_{10}^w < \rho(1 - p_w) - (1 - 2p_w)\rho^2$, and the parameter $f_{10}^w \in \rho p_w(1 \pm \Delta_{10}^w)$. The analysis of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ with respect to f_{10}^w is as follows:

- If $\rho p_w(1 + \Delta_{10}^w) \leq \rho(1 - p_w) - (1 - 2p_w)\rho^2$, the value of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ decreases monotonically as f_{10}^w increases, and hence $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ achieves maximum when $f_{10}^w = \rho p_w(1 - \Delta_{10}^w)$.
- If $\rho p_w(1 + \Delta_{10}^w) > \rho(1 - p_w) - (1 - 2p_w)\rho^2$, the value of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ first decreases and then increases as f_{10}^w increases, and hence $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ achieves maximum when $f_{10}^w = \rho p_w(1 - \Delta_{10}^w)$ or $f_{10}^w = \rho p_w(1 + \Delta_{10}^w)$.

Similarly, the partial derivative of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ with respect to f_{11}^w is given as

$$\begin{aligned} \frac{\partial [I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)]}{\partial f_{11}^w} &= \log \left(\frac{f_{11}^w(1 - f_{10}^w - f_{11}^w)}{(f_{10}^w + f_{11}^w)(f_{*1}^w - f_{11}^w)} \right) + \log \left(\frac{(1 - \rho)(f_{10}^w + f_{11}^w)}{\rho(1 - f_{10}^w - f_{11}^w)} \right) \\ &= \log \left(\frac{f_{11}^w(1 - \rho)}{\rho(f_{*1}^w - f_{11}^w)} \right). \end{aligned} \quad (129)$$

Note that the value of term (129) is positive when $f_{11}^w > \rho p_w + (1 - 2p_w)\rho^2$, and the parameter $f_{11}^w \in \rho(1 - p_w)(1 \pm \Delta_{11}^w)$. The analysis of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ with respect to f_{11}^w is as follows:

- If $\rho(1 - p_w)(1 - \Delta_{11}^w) \geq \rho p_w + (1 - 2p_w)\rho^2$, the value of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ increases monotonically as f_{11}^w increases, and hence $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ achieves maximum when $f_{11}^w = \rho(1 - p_w)(1 + \Delta_{11}^w)$.

- If $\rho(1-p_w)(1-\Delta_{11}^w) < \rho p_w + (1-2p_w)\rho^2$, the value of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ first decreases and then increases as f_{11}^w increases, and hence $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ achieves maximum when $f_{11}^w = \rho(1-p_w)(1+\Delta_{11}^w)$ or $f_{11}^w = \rho(1-p_w)(1-\Delta_{11}^w)$.

It turns out that the maximal value of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ is attained at one of the four “corner” points, *i.e.*, $(f_{10}^w, f_{11}^w) = (\rho p_w(1 \pm \Delta_{10}^w), \rho(1-p_w)(1 \pm \Delta_{11}^w))$. Note that the value of f_{*1}^w has negligible impact since

$$(f_{*1}^w - \rho * p_w) \frac{\partial I(\vec{x}^{(i)}; \vec{y}_w^{(i)})}{\partial f_{*1}^w} \Big|_{\rho * p_w, \rho p_w, \rho(1-p_w)} = \mathcal{O}(n^{-1})$$

while $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ scales as $\mathcal{O}(n^{-1/2})$. Therefore, we set f_{*1}^w to be $\rho * p_w$ for convenience.

APPENDIX E

In Appendix D we have shown that the maximal value of $I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)$ is attained at one of the four “corner” points, *i.e.*, $f_{*1}^w = \rho * p_w$, $f_{10}^w = \rho p_w(1 \pm \Delta_{10}^w)$, $f_{11}^w = \rho(1-p_w)(1 \pm \Delta_{11}^w)$. We now prove that

$$\lim_{n \rightarrow \infty} -k_1 \sqrt{n} \left(I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho) \right) \geq k_1 \max_{i=1}^4 \{g_i(p_w, \epsilon_d, \Delta_{10}^w, \Delta_{11}^w)\}.$$

We first calculate the value of $-k_1 \sqrt{n} \left(I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho) \right)$ at one “corner” point, *i.e.*, when $f_{*1}^w = \rho * p_w$, $f_{10}^w = \rho p_w(1 - \Delta_{10}^w)$, $f_{11}^w = \rho(1-p_w)(1 + \Delta_{11}^w)$, $f_{01}^w = f_{*1}^w - \rho(1-p_w)(1 + \Delta_{11}^w)$, and $f_{00}^w = 1 - f_{*1}^w - \rho p_w(1 - \Delta_{10}^w)$.

$$\begin{aligned} & n^{-k_1 \sqrt{n} [I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)]} \\ &= 2^{n' (f_{01}^w + f_{11}^w) H\left(\frac{f_{11}^w}{f_{01}^w + f_{11}^w}\right) + n' (f_{00}^w + f_{10}^w) H\left(\frac{f_{10}^w}{f_{00}^w + f_{10}^w}\right) + n' (f_{10}^w + f_{11}^w) \log \rho + n' (f_{00}^w + f_{01}^w) \log (1-\rho)} \\ &= 2^{n' (f_{01}^w + f_{11}^w) H\left(\frac{f_{11}^w}{f_{01}^w + f_{11}^w}\right) + n' (f_{00}^w + f_{10}^w) H\left(\frac{f_{10}^w}{f_{00}^w + f_{10}^w}\right) + n' (f_{10}^w + f_{11}^w) \log \rho + n' (f_{00}^w + f_{01}^w) \log (1-\rho)}. \end{aligned} \quad (130)$$

For notational convenience we now consider just the terms in the exponent in (130),

$$\begin{aligned} & n' (f_{01}^w + f_{11}^w) H\left(\frac{f_{11}^w}{f_{01}^w + f_{11}^w}\right) + n' (f_{00}^w + f_{10}^w) H\left(\frac{f_{10}^w}{f_{00}^w + f_{10}^w}\right) + n' (f_{10}^w + f_{11}^w) \log \rho + n' (f_{00}^w + f_{01}^w) \log (1-\rho) \\ &= k_1 \sqrt{n} (\log n) \left[f_{*1}^w H\left(\frac{k_2(1-p_w)(1+\Delta_{11}^w)}{\sqrt{n} f_{*1}^w}\right) + (1-f_{*1}^w) H\left(\frac{k_2 p_w(1-\Delta_{10}^w)}{\sqrt{n}(1-f_{*1}^w)}\right) \right. \\ & \quad \left. + \frac{k_2 p_w(1-\Delta_{10}^w) + k_2(1-p_w)(1+\Delta_{11}^w)}{\sqrt{n}} \log\left(\frac{k_2}{\sqrt{n}}\right) + \frac{\sqrt{n} - (k_2 p_w(1-\Delta_{10}^w) + k_2(1-p_w)(1+\Delta_{11}^w))}{\sqrt{n}} \log\left(\frac{\sqrt{n} - k_2}{\sqrt{n}}\right) \right] \\ &= k_1 \sqrt{n} (\log n) \left[-f_{*1}^w \frac{k_2(1-p_w)(1+\Delta_{11}^w)}{\sqrt{n} f_{*1}^w} \log\left(\frac{k_2(1-p_w)(1+\Delta_{11}^w)}{\sqrt{n} f_{*1}^w}\right) - f_{*1}^w \frac{\sqrt{n} f_{*1}^w - k_2(1-p_w)(1+\Delta_{11}^w)}{\sqrt{n} f_{*1}^w} \log\left(\frac{\sqrt{n} f_{*1}^w - k_2(1-p_w)(1+\Delta_{11}^w)}{\sqrt{n} f_{*1}^w}\right) \right. \\ & \quad \left. - (1-f_{*1}^w) \frac{k_2 p_w(1-\Delta_{10}^w)}{\sqrt{n}(1-f_{*1}^w)} \log\left(\frac{k_2 p_w(1-\Delta_{10}^w)}{\sqrt{n}(1-f_{*1}^w)}\right) - (1-f_{*1}^w) \frac{\sqrt{n}(1-f_{*1}^w) - k_2 p_w(1-\Delta_{10}^w)}{\sqrt{n}(1-f_{*1}^w)} \log\left(\frac{\sqrt{n}(1-f_{*1}^w) - k_2 p_w(1-\Delta_{10}^w)}{\sqrt{n}(1-f_{*1}^w)}\right) \right. \\ & \quad \left. + \frac{k_2 p_w(1-\Delta_{10}^w) + k_2(1-p_w)(1+\Delta_{11}^w)}{\sqrt{n}} \log\left(\frac{k_2}{\sqrt{n}}\right) + \frac{\sqrt{n} - (k_2 p_w(1-\Delta_{10}^w) + k_2(1-p_w)(1+\Delta_{11}^w))}{\sqrt{n}} \log\left(\frac{\sqrt{n} - k_2}{\sqrt{n}}\right) \right] \\ &= k_1 (\log n) \left[k_2(1-p_w)(1+\Delta_{11}^w) \log\left(\frac{\sqrt{n} f_{*1}^w - k_2(1-p_w)(1+\Delta_{11}^w)}{(\sqrt{n} - k_2)(1-p_w)(1+\Delta_{11}^w)}\right) + k_2 p_w(1-\Delta_{10}^w) \log\left(\frac{\sqrt{n}(1-f_{*1}^w) - k_2 p_w(1-\Delta_{10}^w)}{(\sqrt{n} - k_2) p_w(1-\Delta_{10}^w)}\right) \right. \\ & \quad \left. + \sqrt{n} f_{*1}^w \log\left(\frac{\sqrt{n} f_{*1}^w(1-f_{*1}^w) - f_{*1}^w k_2 p_w(1-\Delta_{10}^w)}{\sqrt{n} f_{*1}^w(1-f_{*1}^w) - (1-f_{*1}^w) k_2(1-p_w)(1+\Delta_{11}^w)}\right) + \sqrt{n} \log\left(\frac{(\sqrt{n} - k_2)(1-f_{*1}^w)}{\sqrt{n}(1-f_{*1}^w) - k_2 p_w(1-\Delta_{10}^w)}\right) \right], \end{aligned} \quad (131)$$

As n grows without bound, the term in (131) equals

$$\begin{aligned} & \lim_{n \rightarrow \infty} k_1 (\log n) \left[k_2(1-p_w)(1+\Delta_{11}^w) \log\left(\frac{p_w}{(1-p_w)(1+\Delta_{11}^w)}\right) + k_2 p_w(1-\Delta_{10}^w) \log\left(\frac{1-p_w}{p_w(1-\Delta_{10}^w)}\right) \right. \\ & \quad \left. + k_2((1-p_w)(1+\Delta_{11}^w) - 1) \log e + k_2 p_w(1-\Delta_{10}^w) \log e \right] \\ &= k_1 (\log n) \cdot g_1(p_w, \epsilon_d, \Delta_{10}^w, \Delta_{11}^w), \end{aligned} \quad (132)$$

where the auxiliary multivariable function $g_1(u, v, w, t)$, defined in Section IV, has the form

$$g_1(u, v, w, t) = k_2(u, v) \left[u(1-w) \left(\log\left(\frac{1-u}{u(1-w)}\right) + \log e \right) + (1-u)(1+t) \left(\log\left(\frac{u}{(1-u)(1+t)}\right) + \log e \right) - \log e \right].$$

Similarly, we also calculate the values of $-k_1\sqrt{n} \left(I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho) \right)$ at the other three “corner” points, and it turns out that these values can be characterized by $g_2(p_w, \epsilon_d, \Delta_{10}^w, \Delta_{11}^w)$, $g_3(p_w, \epsilon_d, \Delta_{10}^w, \Delta_{11}^w)$ and $g_4(p_w, \epsilon_d, \Delta_{10}^w, \Delta_{11}^w)$ respectively. It then follows that for sufficiently large n ,

$$\begin{aligned} & \lim_{n \rightarrow \infty} n^{-k_1\sqrt{n}} [I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho)] \\ & \geq 2^{k_1(\log n) \cdot \max_{i=1}^4 g_i(p_w, \epsilon_d, \Delta_{10}^w, \Delta_{11}^w)} \\ & = n^{k_1 \max_{i=1}^4 g_i(p_w, \epsilon_d, \Delta_{10}^w, \Delta_{11}^w)}. \end{aligned}$$

Therefore, we conclude that when the blocklength n is sufficiently large,

$$\lim_{n \rightarrow \infty} -k_1\sqrt{n} \left(I(\vec{x}^{(i)}; \vec{y}_w^{(i)}) + D(\vec{x}^{(i)} \parallel \rho) \right) \geq k_1 \max_{i=1}^4 g_i(p_w, \epsilon_d, \Delta_{10}^w, \Delta_{11}^w).$$

REFERENCES

- [1] M. R. Bloch, “Covert communication over noisy channels: A resolvability perspective,” *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2334–2354, May 2016.
- [2] C. E. Shannon, “Communication theory of secrecy systems,” *Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.
- [3] A. Kerckhoffs, “La cryptographie militaire,” *Journal des Sciences Militaires IX*, vol. 5, no. 38, pp. 161–191, 1883.
- [4] A. D. Wyner, “The wire-tap channel,” *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.
- [5] L. H. Ozarow and A. D. Wyner, “Wire-tap channel II,” *AT&T Bell Laboratories technical journal*, vol. 63, no. 10, pp. 2135–2157, 1984.
- [6] M. R. Bloch and J. N. Laneman, “Strong secrecy from channel resolvability,” *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8077–8098, 2013.
- [7] M. Bloch, J. Barros, M. R. Rodrigues, and S. W. McLaughlin, “Wireless information-theoretic security,” *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2515–2534, 2008.
- [8] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*. Morgan Kaufmann, 2007.
- [9] C. Cachin, “An information-theoretic model for steganography,” *Information and Computation*, vol. 192, no. 1, pp. 41–56, 2004.
- [10] U. M. Maurer, “A unified and generalized treatment of authentication theory,” in *Proceedings of the Annual Symposium on Theoretical Aspects of Computer Science*, pp. 387–398, 1996.
- [11] Y. Wang and P. Moulin, “Perfectly secure steganography: Capacity, error exponents, and code constructions,” *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2706–2722, 2008.
- [12] B. A. Bash, D. Goeckel, and D. Towsley, “Limits of reliable communication with low probability of detection on awgn channels,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, 2013.
- [13] B. Bash, D. Goeckel, and D. Towsley, “LPD communication when the warden does not know when,” in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 606–610, 2014.
- [14] B. A. Bash, A. H. Gheorghe, M. Patel, J. L. Habif, D. Goeckel, D. Towsley, and S. Guha, “Quantum-secure covert communication on bosonic channels,” *Nature communications*, vol. 6, 2015.
- [15] B. A. Bash, D. Goeckel, and D. Towsley, “Square root law for communication with low probability of detection on AWGN channels,” in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 448–452, 2012.
- [16] P. H. Che, M. Bakshi, and S. Jaggi, “Reliable deniable communication: Hiding messages in noise,” in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 2945–2949, 2013, extended version: <http://arxiv.org/abs/1304.6693>.
- [17] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, “Reliable, deniable and hidable communication,” in *Proceedings of the IEEE Information Theory and Applications Workshop*, pp. 1–10, 2014.
- [18] P. Che, M. Bakshi, C. Chan, and S. Jaggi, “Reliable deniable communication with channel uncertainty,” in *Proceedings of the IEEE Information Theory Workshop*, pp. 30–34, 2014.
- [19] P. H. Che, S. Kadhe, M. Bakshi, C. Chan, S. Jaggi, and A. Sprintson, “Reliable, deniable and hidable communication: A quick survey,” in *Proceedings of the IEEE Information Theory Workshop*, pp. 227–231, 2014.
- [20] J. Hou and G. Kramer, “Effective secrecy: Reliability, confusion and stealth,” in *Proceedings of the IEEE Information Theory Workshop*, pp. 601–605, 2014.
- [21] L. Wang, G. W. Wornell, and L. Zheng, “Fundamental limits of communication with low probability of detection,” *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3493–3503, June 2016.
- [22] S. Lee and R. J. Baxley, “Achieving positive rate with undetectable communication over AWGN and Rayleigh channels,” in *Proceedings of the IEEE International Conference on Communications*, pp. 780–785, 2014.
- [23] S. Lee, R. J. Baxley, J. B. McMahon, and R. Scott Frazier, “Achieving positive rate with undetectable communication over MIMO Rayleigh channels,” in *Proceedings of the IEEE 8th Sensor Array and Multichannel Signal Processing Workshop*, pp. 257–260, 2014.
- [24] L. Deshotels, “Inaudible sound as a covert channel in mobile devices,” in *Proceedings of the 8th USENIX Workshop on Offensive Technologies*, pp. 16–16, 2014.
- [25] B. C. Carrara and C. Adams, “On Characterizing and Measuring Out-of-Band Covert Channels,” in *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security*, pp. 43–54, 2015.
- [26] J. Classen, M. Schulz, and M. Hollick, “Practical Covert Channels for WiFi Systems,” in *Proceedings of the IEEE Conference on Communications and Network Security*, pp. 209–217, 2015.
- [27] V. Korzhik, G. Morales-Luna, and M. H. Lee, “On the existence of perfect stegosystems,” *International Workshop on Digital Watermarking*, vol. 192, no. 1, pp. 30–38, 2005.
- [28] B. A. Bash, S. Guha, D. Goeckel, and D. Towsley, “Quantum noise limited optical communication with low probability of detection,” in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 1715–1719, 2013.
- [29] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012.
- [30] R. G. Gallager, “Low-density parity-check codes,” *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 21–28, 1962.
- [31] E. Arikan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.
- [32] I. Csiszar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.
- [33] G. D. Forney Jr, *Concatenated Codes*. M.I.T. Press, Cambridge, MA, 1966.

- [34] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. Springer Science & Business Media, 2006.
- [35] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *The Annals of Mathematical Statistics*, pp. 493–507, 1952.
- [36] W. Feller, *An Introduction to Probability Theory and Its Applications: Volume I*. John Wiley & Sons London-New York-Sydney-Toronto, 1968.